



**Universität
Zürich^{UZH}**

**Professionswissen von Lehrpersonen zu Grundlagen des Leseerwerbs und
Diagnostik bei Leseschwierigkeiten**

Entwicklung und Validierung eines Befragungsinstruments für ausgebildete und
angehende Lehrpersonen und Schulische Heilpädagoginnen und Heilpädagogen

Abhandlung
zur Erlangung der Doktorwürde
der Philosophischen Fakultät
der
Universität Zürich

vorgelegt von
Meike Lietz

Angenommen im Herbstsemester 2020
auf Antrag der Promotionskommission bestehend aus
Prof. Dr. Elisabeth Moser Opitz (hauptverantwortliche Betreuungsperson)
Prof. Dr. Andrea Bertschi-Kaufmann

Zürich, 2021

Zusammenfassung

Leseschwierigkeiten gehören zu den häufigsten Lernproblemen in der Schule, und die Folgen für betreffende Kinder sind gravierend. Lehrpersonen und Schulische Heilpädagoginnen und Heilpädagogen (SHP) übernehmen im Rahmen der Diagnostik von Leseschwierigkeiten eine zentrale Aufgabe. Um diese Aufgabe wahrnehmen zu können, müssen entsprechendes Wissen und Können zu Grundlagen des Leseerwerbs und Diagnostik bei Leseschwierigkeiten vorliegen. In den vergangenen Jahren hat sich ein reges Forschungsinteresse entwickelt, das Professionswissen von Lehrpersonen in den Blick zu nehmen und kognitive Wissensbestände objektiv zu messen. In der vorliegenden Arbeit wurde, ausgehend von bestehenden Modellen zum Professionswissen von Lehrpersonen in verschiedenen Fächern und Domänen, ein Befragungsinstrument entwickelt. Dies wurde an einer Stichprobe ($N = 507$), bestehend aus Laien, Lehrpersonen mit und ohne Berufserfahrung sowie SHP empirisch geprüft. Die Analysen erfolgten auf Grundlage der klassischen Testtheorie und des eindimensionalen Rasch-Modells. Das Instrument wurde im Hinblick auf psychometrische Gütekriterien und die Dimensionalität (u.a. Fachwissen und fachdidaktisches Können) beurteilt. Zudem wurde untersucht, ob sich das Professionswissen in verschiedenen Substichproben (Lehrpersonen und SHP mit unterschiedlicher Berufserfahrung und Ausbildung) unterscheidet und ob das Instrument somit ausbildungssensitiv misst. Die Ergebnisse belegen, dass das Instrument mit Einschränkungen geeignet ist, das Wissen und Können objektiv, reliabel und valide zu messen. Die theoretisch postulierten Dimensionen haben sich nicht wie erwartet gezeigt. Das Instrument ist in der Lage, Unterschiede im Wissen und Können der Substichproben zu erfassen und erfüllt damit das Validitätskriterium der Ausbildungssensitivität.

Abstract

Reading difficulties are one of the most common learning problems in school and the consequences for the children concerned are serious. Special Needs Education Teachers (SHP) play a central role in diagnosing reading difficulties. In order to be able to carry out this task, they must have the knowledge and skills needed to understand the basics of reading acquisition and diagnosis of reading difficulties. In recent years, a lively interest in research has developed in looking at the professional knowledge of teachers and objectively measuring cognitive knowledge. Based on existing models of the professional knowledge of teachers in other subjects and domains, the present study has developed a survey instrument and tested it empirically on a sample consisting of laypersons, teachers with and without professional experience and SHP ($N = 507$). The analyses are based on classical test theory and the one-dimensional Rasch

model. The instrument was assessed in terms of psychometric quality criteria and dimensionality (including content knowledge and pedagogical content knowledge). In addition, it was examined whether professional knowledge differs in different subsamples (teachers and SHP with different professional training and experience) and whether the instrument thus measures educationally sensitive. The results show that the instrument is suitable, with limitations, for measuring knowledge and skills objectively, reliably and validly. The theoretically postulated dimensions did not turn out as expected. The instrument is able to capture differences in the knowledge and skills of the subsamples and thus fulfils the validity criterion of training sensitivity.

Danksagung

Mein besonderer Dank gilt meiner Hauptbetreuungsperson Prof. E. Moser Opitz, die meine Arbeit in den vergangenen Jahren fachlich und persönlich unterstützt hat. Sie gab mir entscheidende Hinweise zur inhaltlichen Ausgestaltung dieser Dissertation. Weiter möchte ich Prof. A. Bertschi-Kaufmann für ihre Unterstützung als Zweitgutachterin danken. Ein spezieller Dank geht an Dr. Erich Ramseier, der mich in statistischen Fragen beraten und unterstützt hat und der durch seine Rückmeldungen massgeblich zum Gelingen der statistischen Auswertungen beigetragen hat. Prof. Dr. Michael Eckhart danke ich dafür, dass er stets dafür gesorgt hat, dass sich meine Aufgabe als Dozentin und die Arbeit an der Dissertation vereinbaren liessen. Der Kommission des Instituts für Forschung, Entwicklung und Evaluation der PHBern möchte ich für die großzügige finanzielle Unterstützung bei der Durchführung und Publikation danken. Meinen Kindern und meinem Mann danke ich für ihre Zeit, das Verständnis und die aufmunternden Worte.

Inhaltsverzeichnis

1 Einleitung.....	9
2 Grundlagen des Lesenlernens und Leseerwerbsprozesse bei Kindern mit Leseschwierigkeiten	17
2.1 Verortung und Eingrenzung der theoretischen Grundlagen.....	17
2.2 Leseschwierigkeit, Leseschwäche, Lesestörung und Legasthenie	18
2.2.1 Verwendete Begriffe.....	18
2.2.2 Definition und Prävalenz nach ICD-10 und DSM-5.....	19
2.3 Komponenten der Lesekompetenz	21
2.3.1 Theoretische Verortung der Lesekompetenz	21
2.3.2 Hierarchieniedrige Prozesse: Basale Lesefertigkeiten und Leseflüssigkeit	24
2.3.3 Hierarchiehohe Prozesse: Sinnverstehendes Lesen	31
2.4 Struktur und Merkmale der deutschen Schriftsprache.....	39
2.5 Vorläuferfertigkeiten des Lesens: Phonologische Informationsverarbeitung	44
2.5.1 Phonologisches Arbeitsgedächtnis	44
2.5.2 Phonologische Bewusstheit.....	45
2.5.3 Benennungsgeschwindigkeit	48
2.6 Leseentwicklung: Entwicklungs- und Stufenmodelle.....	50
2.7 Textschwierigkeit und Leseaufgaben.....	56
2.8 Merkmale der Leseschwierigkeit	60
2.8.1 Schwierigkeiten mit der Phonologischen Informationsverarbeitung	60
2.8.2 Schwierigkeiten beim basalen Lesen.....	61
2.8.3 Schwierigkeiten mit der Leseflüssigkeit.....	65
2.8.4 Schwierigkeiten mit dem Leseverständnis	67
2.9 Diagnostik bei Leseschwierigkeiten.....	72
2.9.1 Diagnostik der phonologischen Informationsverarbeitung	74
2.9.2 Diagnostik der basalen Lesekompetenzen und Leseflüssigkeit	76
2.9.3 Diagnostik des Leseverständnisses	79
2.10 Zusammenfassung und Fazit zu relevanten Wissensgrundlagen im Bereich Lesen und Leseschwierigkeiten	82
3 Professionelle Kompetenz und Professionswissen von Lehrpersonen.....	83
3.1 Theoretische Bestimmung von professioneller Kompetenz und Professionswissen bei Lehrpersonen	83
3.1.1 Kompetenz und Professionswissen	83
3.1.2 Verortung von Professionalität und Professionswissen in der kognitionspsychologischen Expertiseforschung	87

3.2 Kompetenzmodelle und Typologien	90
3.2.1 Typologie professionellen Wissens nach Shulman.....	91
3.2.2 Das Kompetenzmodell der COACTIV-Studie	93
3.3 Operationalisierung des Professionswissens	98
3.3.1 Konzeptualisierung von „ <i>Fachwissen</i> “, „ <i>fachdidaktisches Wissen/Können</i> “ und „ <i>diagnostisches Wissen</i> “	98
3.3.1.1 Fachwissen	98
3.3.1.2 Fachdidaktisches Wissen.....	103
3.3.1.3 Diagnostisches Wissen im Kontext des Professionswissens	107
3.3.2 Zum Verhältnis von Fachwissen und fachdidaktischem Wissen/Können	111
3.3.3 Mentale Repräsentationen von Professionswissen und messmethodische Folgerungen	113
3.3.3.1 Differenzierung von Wissen und Können.....	114
3.3.3.2 Kognitive Anforderungen von Wissen und Können	116
3.4 Wirkung des Professionswissens im Unterrichtskontext	118
3.4.1 Wirkungskette der schulischen Bildung.....	118
3.4.2 Entwicklung von Fachwissen und fachdidaktischem Wissen	119
3.4.3 Einfluss des Fachwissens und fachdidaktischen Wissens auf Schülerinnen- und Schülerleistungen.....	120
3.5 Zusammenfassung und Fazit zum Professionswissen von Lehrpersonen	124
4 Professionswissen von Lehrpersonen im Bereich Grundlagen zum Leseerwerb und Diagnostik bei Leseschwierigkeiten.....	126
4.1 Ableitung des theoretischen Modells des lesebezogenen Wissens/Könnens	126
4.2 Operationalisierung der Inhaltsbereiche Grundlagenwissen und Diagnostisches Wissen.....	130
4.3 Operationalisierung von Fachwissen und fachdidaktischem Wissen/Können unter Einbezug der kognitiven Anforderungen	131
4.3.1 Fachwissen	131
4.3.2 Fachdidaktisches Wissen/Können.....	132
5 Darstellung der Untersuchung und Entwicklung des Befragungsinstruments	135
5.1 Ableitung des Forschungsanliegens und der Forschungsziele.....	135
5.2 Untersuchungsplan	142
5.3 Entwicklung Befragungsinstruments zum lesebezogenen Wissen/Können	143
5.3.1 Beschreibung des Befragungsinstruments.....	143
5.3.2 Entwicklung der Items und Beschreibung der Itemformate.....	146
5.3.3 Qualitative Validierung des Befragungsinstruments	153
5.3.3.1 Befragung von Expertinnen und Experten.....	153
5.3.3.2 Kognitives Pretesting	154
5.3.4 Kodierung der Items und Entwicklung des Kodiermanuals.....	155

5.4 Durchführung der Datenerhebung und Beschreibung der Stichprobe.....	159
5.4.1 Durchführung der Datenerhebung in der Vorstudie und Hauptstudie.....	159
5.4.2 Stichprobe.....	161
6 Statistische Methoden	163
6.1 Testtheoretische Einordnung der Analysen.....	163
6.1.1 Eigenschaften der Item-Response-Theorie.....	163
6.1.2 Vorteile der Item-Response-Theorie gegenüber der Klassischen Testtheorie.....	164
6.1.3 Das Rasch-Modell.....	165
6.2 Kriterien und Analysen zur qualitativen und quantitativen Beurteilung des Befragungsinstruments	169
6.3 Verwendete statistische Verfahren.....	170
6.3.1 Item- und Skalenanalyse in der Klassischen Testtheorie und im Rasch-Modell.....	170
6.3.2 Distraktorenanalyse	175
6.3.3 Differenzielle Itemfunktionsanalyse im Rahmen der Item-Response Theorie.....	175
6.3.4 Dimensionalität des Befragungsinstruments: Modellierung latenter Variablen und deren Zusammenhänge	177
6.3.5 Validierung: Ausbildungssensitivität und Kontrastgruppenvergleich.....	180
7 Ergebnisse.....	183
7.1 Ergebnisse der Voruntersuchung	183
7.1.1 Item- und Skalenanalyse	183
7.1.2 Dimensionalität des Befragungsinstruments	186
7.1.3 Validierung des Befragungsinstruments.....	188
7.2 Ergebnisse der Hauptstudie.....	189
7.2.1 Item- und Skalenanalyse	189
7.2.1.1 Itemtrennschärfe, Itemvarianz und Reliabilität in der Klassischen Testtheorie.....	189
7.2.1.2 Gütekriterien und Skalierung auf Grundlage des Rasch-Modells	191
7.2.1.3 Itemschwierigkeitsanalyse.....	194
7.2.1.4 Fazit aus der Item-Skalenanalyse	200
7.2.2 Ergebnisse der Distraktorenanalyse	201
7.2.3 Ergebnisse der differenziellen Itemfunktionsanalyse (DIF)	203
7.2.3.1 Graphische Inspektion der differenziellen Itemfunktion.....	203
7.2.3.2 Statistische Analyse der differenziellen Itemfunktion	205
7.2.3.3 Interpretation der differenziellen Itemfunktion.....	207
7.2.3.4 Splitten von Items.....	210
7.2.3.5 Fazit aus der differenziellen Itemfunktionsanalyse.....	212
7.2.4 Ergebnisse zur Dimensionalität des Befragungsinstruments	213
7.2.4.1 Dimensionierung des Professionswissens nach Inhalten.....	213

7.2.4.2 Dimensionierung des Professionswissens nach kognitiven Anforderungen	214
7.2.4.3 Exploratorische Faktorenanalyse	215
7.2.4.4 Inhaltliche Beschreibung der ermittelten Faktoren	219
7.2.4.5 Fazit zur Dimensionalisierung des Befragungsinstruments.....	221
7.2.5 Validierung: Ausbildungssensitivität und Kontrastgruppenvergleich.....	221
7.2.5.1 Haupteffekte des Vergleichs der drei Substichproben und der Kontraststichprobe	222
7.2.5.2 Einzelvergleich der drei Substichproben und Kontraststichprobe	223
7.2.5.3 Fazit zur kriterienorientierten Validität	224
7.3 Die Ergebnisse der Hauptstudie im Überblick	225
7.3.1 Zusammenfassung der Distraktorenanalyse, Item-Skalenanalyse und Itemfunktionsanalyse	225
7.3.2 Zusammenfassung der Ergebnisse zur Dimensionierung des Befragungsinstruments	227
7.3.3 Zusammenfassung der Ergebnisse des Kontrastgruppenvergleichs und zur Ausbildungs- sensitivität.....	229
8 Diskussion	231
8.1 Diskussion des Befragungsinstruments	231
8.1.1 Objektivität des Befragungsinstruments und ökonomische Durchführung	231
8.1.2 Psychometrische Kriterien des Befragungsinstruments	236
8.1.3 Struktur des Professionswissens.....	239
8.1.4 Validität des Befragungsinstrument	242
8.1.5 Validität des Befragungsinstruments im Rahmen von Gruppenvergleichen	245
8.2 Limitation des Befragungsinstruments.....	252
8.2.1 Theoretische Aspekte.....	252
8.2.2 Methodische Aspekte.....	254
8.3 Ertrag für Forschung und Praxis	256
9 Literaturverzeichnis.....	259
10 Abbildungsverzeichnis	283
11 Tabellenverzeichnis.....	285
12 Anhang.....	286

1 Einleitung

Zahlreiche nationale und internationale Bildungsmonitorings haben sich in den letzten Jahren mit den Leseleistungen von Kindern und Jugendlichen befasst. Bekannte Studien waren insbesondere internationale Rankings wie die PISA-Studie (Konsortium PISA, 2018), die PIRLS-Studie (Mullis, Martin, Foy & Drucker, 2012) bzw. IGLU¹-Studie (Hußmann et al., 2017) sowie die auf Deutschland bezogene Bundesländer-Vergleichsstudie VERA (Hosenfeld & Zimmer-Müller, 2014). Insbesondere die wiederkehrend publizierten Ergebnisse der PISA-Studien erfuhren medial große Resonanz und führten vor Augen, dass die Leseleistungen in Deutschland und der Schweiz nicht auf dem gewünschtem Niveau liegen (Pissarek & Schilcher, 2017). Schwierigkeiten im Lesen (und auch im Rechtschreiben) gehören gemäß Hartmann (2013) zu den häufigsten Lernproblemen. In der Schweiz beenden nach aktuellster PISA-Daten im Jahr 2015 20% der Schülerinnen und Schüler die obligatorische Schulzeit mit Lesefähigkeiten, die unter dem Kompetenzniveau 2 liegen (Konsortium PISA, 2018). Das heißt, dass sie nur sehr einfache Leseaufgaben in angeleiteten Kontexten bewältigen können (Nix, 2010). Dies ist insofern alarmierend, da dies für eine Vielzahl von sich in Ausbildungs- und Alltagssituationen stellenden Leseanforderungen nicht hinreichend ist (Naumann, Artelt, Schneider & Stanat, 2010).

Kinder, die kaum lesen können, sind in vielerlei Hinsicht benachteiligt, wobei man von einer Kette von Folgen für Ausbildungs-, Berufs- und Alltagssituationen sprechen kann, die ihren Anfang in der schulischen Laufbahn nehmen (vgl. u.a. Kemmler, 1976; Naumann et al., 2010; Sturm, 2012; Wember, 1999; Wiesner & Schneider, 2011). Die überwiegende Mehrheit schulversagender Kinder sind Schülerinnen und Schüler mit äußerst schwachen Lese- und Rechtschreibleistungen (Wember, 1999). Da die Lesefähigkeit eine Grundkompetenz zur Bewältigung der alltäglichen Informationsvielfalt ist und nicht zuletzt auch „Mitgliedschaft“, das heißt Zugehörigkeit zu unserer Gesellschaft bedeutet (Bertschi-Kaufmann, 2011, S. 8), ist der bestehende Anteil an Jugendlichen, welche die Schule mit unzureichenden Lesekompetenzen verlassen, auch in dieser Hinsicht problematisch.

Der Schule kommt im Bereich der Leseförderung eine enorme Bedeutung zu, weshalb die PISA-Studien in der Schweiz wie auch in Deutschland einen „enormen Schub an Forschung und Entwicklung zum Lesen und zum Leseunterricht ausgelöst“ haben (Schneider & Lindauer, 2011, S. 1). Von Lehrpersonen wird ein vertieftes Wissen zu Prozessen des Lesenlernens und zur Diagnostik bei Leseschwierigkeiten verlangt (Artelt, 2009; Behrmann & Souvignier, 2013).

¹ IGLU ist die in Deutschland durchgeführte Teilstudie von PIRLS.

Gefordert werden verstärkte Anstrengungen in der Aus- und Fortbildung von Lehrpersonen, damit die fachlichen und fachdidaktischen Kompetenzen² gestärkt werden können, denn „wer etwas an der Schule verbessern möchte, sollte die Lehrkräfte in den Blick nehmen“ (König & Hofmann, 2010, S. 9). Dieses Bestreben ist auch vor dem Hintergrund des festgestellten Wissensdefizits bei Lehrpersonen zu empirisch fundierten Grundlagen und Fördermethoden bei Leserechtschreibschwierigkeiten (LRS) nachvollziehbar, ein Defizit, das in mehreren Studien bestätigt wird (Hintz & Grünke, 2009; Schmidt & Schabmann, 2016; Sodogé, 2010; Steck, 2010). Insbesondere zeigen sich in verschiedenen Untersuchungen eher gering ausgeprägte diagnostische Kompetenzen im Bereich Leseschwierigkeiten, was sich darin manifestiert, dass leseschwache Kindern oft gar nicht erkannt werden und deren Leistungen tendenziell überschätzt werden (Helmke, 2012; Karing, Matthäi & Artelt, 2011; Rjosk, McElvany, Anders & Becker, 2011; Schmidt & Schabmann, 2010).

Genau dieses „in den Blick nehmen der Lehrpersonen“ geschieht seit einigen Jahren im Rahmen der Forschung zur professionellen Kompetenz von Lehrpersonen (Baumert & Kunter, 2011a). Lehrpersonen als Expertinnen und Experten von Unterrichtsgestaltung rücken vermehrt in den Mittelpunkt, wie Baer et al. (2011) feststellen. Einen wichtigen Impuls für aktuelle Ansätze der Professionsforschung gaben Baumert und Kunter (2006, S. 482) mit ihrem „Modell professioneller Handlungskompetenz“, das in seiner Grundlegung an die Typologie von Shulman (1986) anschließt. Neben distalen Faktoren wie Bildungsabschlüssen oder besuchten Weiterbildungen befassen sich Baumert & Kunter (2006) mit der Frage, wie handlungsnähere, auf das pädagogisch-didaktische Wissen und Können bezogene Faktoren messbar gemacht werden können. Gemäß dem Autor und der Autorin liegt das eigentliche Forschungsdesiderat bei der direkten, reliablen und validen Erfassung von Lehrerinnen- und Lehrerkompetenzen, wobei die besondere Herausforderung darin besteht, Kombinationen von Wissen und praktischem Können zu erfassen und zudem eine separierte Indikatorisierung von Fachwissen und fachdidaktischem Wissen vorzunehmen (Baumert & Kunter, 2006). Für die theoretische Konzeptualisierung ist eine Orientierung an der Expertiseforschung erfolgversprechend (Baumert & Kunter, 2006), bei der – wie auch bei anderen Forschungsparadigmen (Persönlichkeitsparadigma, Prozess-Produkt Paradigma) – die Frage nach „dem guten Lehrer“ im Zentrum steht. Der Fokus wird bei diesem Zugang auf das kognitiv abrufbare Wissen und Können gelegt (Krauss & Bruckmaier, 2014), weshalb dieser Forschungsansatz auch der Lehrer- bzw. Lehrerinnen-

² In dieser Arbeit wird teilweise der Begriff „Kompetenz“ gemäß zitierter Quelle verwendet, im Wissen darum, dass Kompetenz weiter gefasst ist als das Professionswissen. Eine Abgrenzung von „Kompetenz“ und „Professionswissen“, wie sie in der vorliegenden Arbeit verstanden wird, findet sich in Kap. 3.1.1.

kognitionsforschung zugeordnet wird. Im Bereich der quantitativen Empirie spricht Tillmann (2014, S. 312) im Zusammenhang mit dem Expertiseansatz von einer „Leitfigur der Lehrerforschung“.

Stand der Forschung

Ein Blick auf den Stand der Forschung zeigt, dass sich in den letzten Jahren im Bereich der empirischen Messung von kognitivem, fachbezogenem Professionswissen ein reges Interesse entwickelt hat und ein rasanter Anstieg an Publikationen zu verzeichnen ist (siehe Überblick bei Krauss, Lindl, Schilcher & Tepner, 2017b). Es geht um das Interesse, Kompetenzen von Lehrpersonen zu systematisieren, operationalisieren und den Einfluss auf die Leistungen der Schülerinnen und Schüler festzustellen. Begründet wird dieses Interesse damit, dass die Kompetenz der Lehrpersonen einen relevanten Einfluss auf die Schülerinnen- und Schülerleistungen hat (vgl. u.a. Hattie, 2003; Lipowsky, 2006). Dies gilt zumindest dann, wenn der Blick auf *veränderbare* Bedingungsfaktoren (fachliche Kompetenzen) gelenkt wird (König, 2010). Diese Aussage hat mit Bezug auf neuere Publikationen nicht an Aktualität verloren, und die besondere Bedeutung des fachdidaktischen Wissens von Lehrpersonen für die Unterrichtsqualität und Leistungsentwicklung der Schülerinnen und Schüler ist inzwischen mehrfach belegt (König et al., 2018; Schlicher & Krauss, 2019).

Im Bereich empirischer Messungen fachwissenschaftlicher und fachdidaktischer Kompetenz von (angehenden) Mathematiklehrpersonen der Primar- und/oder Sekundarstufe I sind in den letzten Jahren Large Scale-Studien entstanden, z.B. die COATIV-Studie [*Cognitive Activation in the Classroom*] (Kunter et al., 2011a) oder die internationale Vergleichsstudie MT21 [*Mathematics Teaching in the 21st Century*] (Blömeke, Kaiser & Lehmann, 2008), die als Vorbereitung auf die TEDS-M Studie [*Teacher Education and Development Study – Mathematics*] (Blömeke, Kaiser & Lehmann, 2010a/2010b) verstanden wurde. Zeitlich vorangehend wurden an der University of Michigan (Hill, Schilling & Loewenberg Ball, 2004) entsprechende Messinstrumente zur Erfassung der Fachkompetenz von Grundschullehrpersonen im Fach Mathematik empirisch geprüft. Eine Ausweitung auf andere Fächer (Deutsch, Englisch, Latein, Physik, Musik und Religion, Naturwissenschaften, Geschichte, Technik) fand etwas später statt, und zwar im Rahmen der TEDS-LT [*Teacher Education and Development Study – Learning to Teach*] (Blömeke, 2011; Bremerich-Vos, Dämmer, Willenberg & Schwippert, 2011) sowie der FALKO-Studie [*Fachspezifische Lehrerkompetenzen*] (Krauss et al., 2017a) und der EKoL-Studie [*Effektive Kompetenzdiagnose in der Lehrerbildung*] (Dörfler, Schmitt, & Rutsch, 2015) statt. Bei aller Unterschiedlichkeit weisen die aufgeführten Studien bezüglich

„Konzeptualisierung des Konstrukts «professionelle Kompetenz» sowie in den damit verbundenen grundlegenden Prämissen große Gemeinsamkeiten auf“ (König, 2010, S. 50). Dies betrifft u.a. die Verwendung der Erkenntnisse aus der Expertiseforschung, die Grundannahmen zur Erfassung von Wissen und Können sowie übergreifende Modelle professioneller Kompetenz von Lehrpersonen (König, 2010). Insbesondere erfolgte eine getrennte Konzeptualisierung der drei Wissensbereiche *Fachwissen*, *fachdidaktisches Wissen* und *pädagogisches Wissen* (Krauss et al., 2017b). Die vorgestellten Studien markieren so etwas wie den „state of the art“ im Zusammenhang mit der Testkonstruktion zum Professionswissen im Lehrberuf. Die vorliegende Studie orientiert sich insbesondere methodisch, aber auch bezüglich der formalen Operationalisierung von Fachwissen und fachdidaktischem Wissen/Können an den erwähnten Studien.

Der Forschungsstand zu kognitiven Wissensbeständen von Lehrpersonen im Bereich *Lesen* ist relativ überschaubar und wird im Rahmen verschiedener Forschungsüberblicke im Zusammenhang deutschdidaktischer Lehrforschung nachgezeichnet (siehe Überblicke bei Lessing-Sattari & Wieser, 2018; Pissarek & Schilcher, 2017; Wiprächtiger-Geppert, Riegler & Freivogel, 2015). Unter Einbezug der in den Überblicken genannten Studien und weiterer recherchierter Untersuchungen zeichnet sich folgendes Bild ab: Zum einen liegen qualitative Studien vor, in denen handlungsleitende Kognitionen zum Leseunterricht (Scherf, 2013; Schmill, 2013) und fachdidaktische Konzepte zum Literaturunterricht (Wieser, 2008) sowie das lesebezogene praktische professionelle Wissen von Lehrpersonen (Bräuer, 2010) erfasst wurden. Weiter steht in der Fallstudie von Steck (2010) die Sach-, Diagnose- und Förderkompetenz von Lehrpersonen im Bereich Leseverstehen im Fokus. In der qualitativen Studie von Kamzela (2015) wurde u.a. der Frage nach den Konzepten von basalem Lesen und Leseunterricht Lehrpersonen nachgegangen. Und schließlich nimmt die Studie von Schmidt (2015) Vorstellungen und Orientierungen von Lehrpersonen zur Diagnose von Lesekompetenz in den Blick. In einer Studie von Herzmann (2010) wurden zudem Zusammenhänge zwischen dokumentierten Unterrichtseinheiten und Interview-Aussagen von Lehrpersonen sowie Schülerinnen und Schülern inhaltsanalytisch ausgewertet und mit deren Lesekompetenzen verbunden. Alle erwähnten Studien sind im Bereich qualitativer Fallstudien anzusiedeln, in deren Rahmen mittels relativ kleiner Stichproben (3 - 21 Lehrpersonen) domänenspezifische Wissensbestände und Überzeugungen erfasst wurden.

Neben qualitativen Studien liegen einige Fragebogenuntersuchungen vor, in denen über Selbstauskünfte verschiedene Aspekte professionellen Wissens von Lehrpersonen im Bereich Leseunterricht und spezifisches Wissen zu Lese-Rechtschreibungsschwierigkeiten erfasst wurde (Hefti

& Ventocilla Metz, 2009; Inckemann, 2008; Sodogé, 2010). Zwei weitere Studien, die sich auf Selbstauskünfte von Lehrpersonen stützen, liegen von Isler und Leemann (2008) sowie Basilicata und Meier (2011) vor. Sie fokussieren auf die didaktische Gestaltung und die fachdidaktischen Konzepte von Lehrpersonen bezüglich deren Lese- und/oder Schreibförderung. In den genannten Studien wurden auf verschiedenen Ebenen Wissensbestände aber auch subjektive Theorien und Auskünfte über unterrichtliches Handeln erfasst. Zum einen fällt auf, dass bis dahin nur vereinzelt (z.B. Scherf, 2013; Wieser, 2008) Bezüge zu Kompetenzmodellen bzw. theoretischen Taxonomien des Professionswissens hergestellt wurden, was damit zusammenhängen dürfte, dass Systematisierungen professioneller Kompetenz erst mit den genannten Large-Scale Studien überhaupt populär wurden. Zum anderen zeigt sich, dass die Entwicklung psychometrischer Tests für Lehrpersonen im engeren Sinne erst in neusten Studien Eingang gefunden hat – angestoßen durch die COACTIV-Studie (Krauss et al., 2017b). Im Bereich Lesen sind hier die zeitgleich zum vorliegenden Projekt entstandenen Messinstrumente von Rutsch (2016), Schmidt und Schabmann (2016) und Hanke et al. (2019) zu erwähnen. In allen Untersuchungen wurden Tests entwickelt, die auf die objektive, quantifizierbare Erfassung von Wissen im Bereich fachdidaktischen Wissens zum Leseunterricht (Rutsch, 2016), Wissen zum Umgang mit Lese-Rechtschreibschwierigkeiten (LRS) (Schmidt & Schabmann, 2016) sowie Wissen zum basalen Lesen- und Schreibenlernen (Hanke et al., 2019) zielten. In den Studien wurde auch die Berufsgruppe der Sonderschullehrpersonen³ einbezogen.

Für die vorliegende Studie waren aufgrund der inhaltlichen Nähe von Schreib- und Leseerwerb (Prozessmodelle, Entwicklungsstufen) auch Studien aus dem Bereich Schriftspracherwerb gewinnbringend, welche die *schreibbezogene* Seite des Schriftspracherwerbs auf Primarstufe in den Blick nahmen. In verschiedenen Untersuchungen wurden an kognitionspsychologischen Grundsätzen der Expertiseforschung orientierte, objektive Messinstrumente entwickelt, die an der Konzeptualisierung von Fachwissen und fachdidaktischem Wissen vorangegangener Studien anknüpften und für die die Domänen⁴ „Orthographie“ (Corvacho del Toro, 2013; Jagemann, 2016; Riegler & Wiprächtiger-Geppert, 2016) und „Schreiben“ spezifiziert wurden. Mit Blick auf den amerikanischen Sprachraum lässt sich der Stand der Forschung zur objektiven Messung von Professionswissen in der Domäne Lesen ergänzen. Angeführt wurde die

³ Die in Deutschland übliche Berufsbezeichnung "Sonderschullehrperson" und die in der Schweiz übliche Bezeichnung "Schulische Heilpädagogin/Schulischer Heilpädagoge" (Akronym SHP) werden synonym verwendet.

⁴ Der Begriff „Domäne“ wird an einigen Stellen praktisch gleichbedeutend mit dem Begriff „Fach“ verwendet (z.B. TEDS-LTBlömeke et al., 2011). An anderen Stellen wird er als spezifischer Teilbereich eines Wissensgebiets verwendet (z.B. "Orthographie" Riegler & Wiprächtiger-Geppert, 2016). In Analogie dazu kann die „Domäne Lesen“ verstanden werden als umgrenztes Teilgebiet des Faches Deutsch und konkreter als das leseseitige Wissensgebiet des Schriftspracherwerbs.

Entwicklung von Messinstrumenten im Bereich des Schriftspracherwerbs und des Lesens von Moats (1994), die einen Fragebogen zum linguistischen Wissen entwarf. Das stark linguistisch ausgerichtete Instrument wurde später von Moats & Foorman (2003) weiterentwickelt und in adaptierter, unterrichtsbezogener Form in weiteren Studien zum Einfluss des Wissens von Lehrpersonen auf die Schülerinnen- und Schülerleistungen eingesetzt (Carlisle, Correnti, Phelps & Zeng, 2009; Carlisle, Kelcey, Rowan & Phelps, 2011; McCutchen et al., 2002). Etwa zeitgleich entstand das Instrument von Phelps & Schilling (2004) zur Erfassung des fachlichen und fachdidaktischen Wissens von Lehrpersonen im Bereich Erstleseunterricht. Schließlich entwickelten Binks-Cantrell, Joshi und Washburn (2012) einen Fragebogen, der das Basiswissen von Lehrpersonen für eine erfolgreiche Leseinstruktion im Anfangsunterricht misst. Somit lassen sich trotz linguistischer Unterschiede zwischen Deutsch und Englisch auch aus Studien aus dem amerikanischen Sprachraum Hinweise zur Operationalisierung des hier interessierenden Konstrukts finden (Kap. 3.3.1).

Bisherige Messungen professioneller Kompetenz beziehen sich vorwiegend auf den Regelschulbereich, während Untersuchungen zum fachspezifischen professionellen Wissen von Sonderschullehrpersonen mit wenigen Ausnahmen fehlen (Jandl & Moser Opitz, 2017). Die Diskussion um das „sonderpädagogische Professionswissen“ wurde bisher eher auf theoretischer Ebene geführt. Zudem liegen Ausbildungsmodelle sowie Beschreibungen von Anforderungsprofilen vor (Jandl & Moser Opitz, 2017). Auf eher unspezifischer Ebene zeigt sich in zahlreichen Bestimmungsversuchen (Benkmann, 2005; Joller-Graf & Sturny-Bossart, 2010; Melzer & Hillenbrand, 2013; Melzer, Hillenbrand, Sprenger & Hennemann, 2015), wie sich sonderpädagogisches Professionswissen zusammensetzt und sich vom Professionswissen von Regellehrpersonen abgrenzt. In Studien der Zürcher-Gruppe um Moser Opitz wurde erstmals der Fokus auf das Professionswissen von Sonderpädagoginnen und Sonderpädagogen gelegt. Kognitive Wissensbestände dieser Berufsgruppe wurden vor dem Hintergrund aktuell diskutierter Kompetenzmodellen systematisiert und empirisch zugänglich gemacht (Jandl & Moser Opitz, 2017), beispielsweise im Rahmen einer Studie zu fachspezifischen professionellen Kompetenzen von Sonderpädagoginnen und Sonderpädagogen auf Kindergartenstufe (Seemann et al., 2017). Zudem wurden entsprechende Instrumente zum Professionswissen im Bereich Mathematik entwickelt (Hepberger, Lindmeier, Moser Opitz & Heinze, 2017; Jandl, 2016), womit eine wichtige Forschungslücke geschlossen wurde. Da Sonderpädagoginnen und Sonderpädagogen aufgrund ihrer Zusatzausbildung bei der Begleitung von Kindern mit besonderem Förderbedarf meist spezifische Aufgaben übernehmen, ist die Frage nach deren Kompetenzen von besonderem Interesse – sowohl für die Mathematik wie auch für den Bereich der Sprache.

Forschungsdesiderat und Vorgehen

Ein noch zu bearbeitendes Forschungsfeld liegt in der Messung des Professionswissens von Lehrpersonen und SHP in der Domäne *Lesen* auf der Primarstufe. Insgesamt weisen Forschungsergebnisse darauf hin, dass das fachliche Wissen von SHP in verschiedenen Bereichen bedeutsam zu sein scheint (Pool Maag & Moser Opitz, 2014). Die Besonderheit des vorliegenden Projektes liegt in der *Mit*-Berücksichtigung des Wissens von Sonderpädagoginnen und Sonderpädagogen, denen aufgrund ihrer Ausbildung bei der Prävention und Unterstützung leseschwacher Kinder eine besondere Aufgabe zukommt (Hartmann, 2007). Vor dem aktuellen Stand der Forschung scheint es zudem dringlich, der Forderung des Einbezugs anschlussfähiger Kompetenzmodelle, die relevante Wissensbereiche unterscheiden, nachzukommen. Ziel der vorgestellten Studie ist die Entwicklung eines Instruments, mit dem das professionelle Wissen (angehender) Lehrpersonen und SHP in verschiedenen Phasen der Ausbildung in den Bereichen „Grundlagen des Leseerwerbs“ und „Diagnostik bei Leseschwierigkeiten“, erfasst werden kann. Der Bereich der *Förderung* von Kindern mit Leseschwierigkeiten wird in der vorliegenden Forschungsarbeit – und entsprechend auch in den theoretischen Grundlagen – nur am Rande behandelt. Die Entscheidung, den Bereich der „Förderung“ nicht stärker zu gewichten, erfolgte aufgrund der Einschätzung, dass es kaum möglich ist, ein Befragungsinstrument zu entwickeln, das ökonomisch und gleichzeitig in der Lage ist, alle Bereiche zur Domäne des Lesens in angemessener Tiefe zu erfassen.

Eine Beschreibung der Forschungsanliegen im Einzelnen findet sich in Kap. 5.1 und ergibt sich aus den im folgenden Kapitel dargestellten theoretischen und empirischen Grundlagen des Leseerwerbs und des Leseerwerbs bei Leseschwierigkeiten (Kap. 2) sowie aus den Grundlagen zur Professionellen Kompetenz und zum Professionswissen (Kap. 3). Für das entwickelte Befragungsinstrument wird der Fokus ausschließlich auf das Lesen gerichtet, im Wissen darum, dass viele leseschwache Kinder auch Rechtschreibschwierigkeiten aufweisen (Moll & Landerl, 2011). In Kap. 4 wird auf der Grundlage der bearbeiteten theoretischen Erkenntnisse das Modell des lesebezogenen Wissens/Könnens abgeleitet, das den Hintergrund des Befragungsinstruments bildet.

Das entwickelte Befragungsinstrument wurde in einer empirischen Studie (Kap. 5 - Kap. 7), die sich in eine Vor- und Hauptstudie gliedert, hinsichtlich psychometrischer Gütekriterien, Objektivität, Reliabilität und Validität beurteilt. In der anschließenden Diskussion (Kap. 8) werden die Möglichkeiten und Grenzen des Befragungsinstruments beschrieben, und die Frage nach den quantitativ feststellbaren Unterschieden des Professionswissens zwischen

Lehrpersonen mit verschiedener Berufsbiografie (Ausbildung/Berufserfahrung) wird nochmals aufgegriffen.

2 Grundlagen des Lesenlernens und Leseerwerbsprozesse bei Kindern mit Leseschwierigkeiten

2.1 Verortung und Eingrenzung der theoretischen Grundlagen

Der Fokus der vorliegenden Studie liegt auf theoretischen Wissensgrundlagen, die Lehrpersonen und SHP für eine erfolgreiche Begleitung zum Erwerb und Aufbau von Lesekompetenzen bis etwa Ende der vierten Klasse besitzen müssten. Dies ist damit begründet, dass in der Regel etwa im Alter von neun Jahren (vierte Klasse) „die Automatisierung beim Erkennen von Wörtern und Sätzen soweit fortgeschritten [ist], dass die Kinder Texte lesen können, die ihren intellektuellen Bedürfnissen entsprechen“ (Rosebrock, 2012, S. 3).

Es zeigt sich, dass sich unterschiedliche Fertigkeiten, welche die spätere Leseentwicklung voraussagen, bereits vor der Einschulung und später im Erstleseunterricht sowie im weiterführenden Leseunterricht zu beobachten sind (Klicpera, Schabmann & Gasteiger-Klicpera, 2013). Die Eingrenzung auf den Zeitraum bis zur vierten Klasse ist damit begründet, dass Kinder mit massiven Leseschwierigkeiten oft in der dritten und vierten Klasse identifiziert werden (Günther, 2007), weshalb das diagnostische Wissen der Lehrpersonen auf dieser Schulstufe besonderes wichtig ist. Anschlusskommunikation, Lesesozialisation und literarische Bildung (siehe Modell Rosebrock & Nix, 2012, S. 92) werden in der vorliegenden Arbeit nur am Rande besprochen, im Wissen darum, dass diese Aspekte die Domäne des Lesens erst vervollständigen.

Der Leseerwerb in einem umfassenden Verständnis ist Ende der vierten Klasse zwar bei weitem nicht abgeschlossen. Der Fokus wird jedoch ab dem vierten Schuljahr mit zunehmend komplexeren Lese- und Verstehensprozessen meist verstärkt auf den „Literaturunterricht“ gelenkt, wodurch sich neue Aufgaben stellen (McElvany & Schneider, 2009, S. 157). Die basalen Lesefertigkeiten werden bei unauffällig lesenden Kindern in der Regel in diesem Zeitraum gefestigt sein, und es kommt im Normalfall zu einer Verlagerung von Sinn und Zweck des Lesens. Die Phase des „learning to read“ wird von der Phase des „reading to learn“ abgelöst (Rosebrock & Nix, 2012, S. 75).

Da der Anspruch des Befragungsinstruments auch darin besteht, das diagnostische Wissen in Bezug auf Kinder mit Leseschwierigkeiten abzufragen, wird in den folgenden theoretischen Grundlagen auch Bezug auf den „beeinträchtigten“ Leseerwerbsprozess genommen, und die Merkmale von Kindern mit Leseschwierigkeiten werden beschrieben.

2.2 Leseschwierigkeit, Leseschwäche, Lesestörung und Legasthenie

2.2.1 Verwendete Begriffe

Es gibt in der Fachliteratur keine einheitliche Begriffsverwendung für Kinder, die Schwierigkeiten beim Lesenlernen und unterdurchschnittliche Leseleistungen haben. Mit Begriffen wie „Lesestörung“, „Leseschwierigkeit“ und „Legasthenie“ sind zum Teil unterschiedliche Konzepte verbunden, die mit verschiedenen Annahmen zur Symptomatik und den Ursachen begründet werden (Schulte-Körne, 2014a, S. 139). Teilweise werden die Bezeichnungen in der Fachliteratur aber auch synonym verwendet (Günther, 2007), und es ist davon auszugehen, dass sich z.B. die Begriffe „Lese-Rechtschreib-Schwäche“ und „Lese-Rechtschreib-Schwierigkeiten“ nicht in ihrer Definition, aber in ihrer Konnotation unterscheiden (Scheerer-Neumann, 2018, S. 18). *Schwierigkeit* fokussiert auf aktuelle Probleme im Lernprozess, während *Schwäche* stärker mit Eigenschaften des Kindes verbunden wird (ebd.). Beim Begriff der *Schwäche* werden Umweltfaktoren mit einbezogen, die mit Dispositionen des Kindes interagieren.

Der Begriff „*Legasthenie*“ ist meist mit einer genetisch verursachten, ausgeprägten Lese- und/oder Rechtschreibstörung assoziiert, während der Begriff der Leseschwäche auch für vorübergehende Schwierigkeiten im Lesen genutzt wird und damit vom Begriff der Lesestörung abgegrenzt wird (Schulte-Körne, 2014a, S. 139). Als „*Lesestörung*“ wird eine anhaltende, bedeutsame Schwäche im Erlernen des Lesens bezeichnet, die weder auf das Entwicklungsalter noch auf eine unzureichende Beschulung, widrige psychosoziale Umstände oder körperliche Ursachen zurückzuführen ist (Schulte-Körne & Galuschka, 2019, S. 1). Abgrenzungsversuche verschiedener Begrifflichkeiten, die u.a. mit Annahmen über verschiedene Ursachen und einer mehr oder weniger starken Fokussierung auf umwelt- und kindbezogene Faktoren zusammenhängen, haben zu keiner einheitlichen Begriffsverwendung geführt (Bundesverband Legasthenie und Dyskalkulie, 2018). In der vorliegenden Arbeit wird der Begriff der *Leseschwierigkeit* in Anlehnung an Günther (2007) im Sinne einer „Sammelkategorie“ (S. 65) verwendet und umfasst alle Schwierigkeiten, die Kinder beim Erlernen des Lesens haben. „*Leseschwierigkeit*“ bezeichnet vorübergehende oder anhaltende Probleme beim Erlernen des Lesens, die sich im Prozess der basalen Lesefertigkeiten, der Leseflüssigkeit und des Leseverständnisses manifestieren (siehe Merkmale der Leseschwierigkeit, Kap. 2.8).

2.2.2 Definition und Prävalenz nach ICD-10 und DSM-5

Die „*International Classification of Diseases*“ (ICD-10)⁵ der Weltgesundheitsorganisation (WHO) und das „*Diagnostic and Statistical Manual of Mental Disorders*“ (DSM-5) von der American Psychiatric Association (2013) definieren die „Lesestörung“⁶ im Sinne eines internationalen Konsenses (Scheerer-Neumann, 2018, S. 24ff.). Deutsche Übersetzungen liegen für die ICD-10 von Dilling, Mombour, Schmidt und Schulte-Markwort (2011) und für das DSM-5 von Falkai und Wittchen (2015) vor.

Die Lesestörung wird in der ICD-10 als Störung umschrieben, bei der insbesondere eine Beeinträchtigung der Lesefertigkeiten und des Leseverständnisses beobachtet wird, d.h. „die Fähigkeit, gelesene Worte wiederzuerkennen, vorzulesen und die Leistungen bei Aufgaben, für welche Lesefähigkeit benötigt wird, können alle betroffen sein“ (Dilling et al., 2011, S. 334). Es handelt sich also um eine Schwäche, die zu erheblichen Defiziten führt, ausdrücklich vorgegebene Informationen zu finden, die Informationen aufeinander zu beziehen und vorwissensbasierte Schlussfolgerungen zu ziehen (Philipp, 2012a). Es wird davon ausgegangen, dass die Störung oft persistierend bis in die Adoleszenz anhält, auch wenn einige Fortschritte gemacht werden (Dilling et al., 2011). Zudem müssen die Leseleistungen unter dem Niveau liegen, das aufgrund von Alter, Intelligenz und Schulung zu erwarten ist. Die Leseschwierigkeit wird durch individuell angewendete, standardisierte Tests zu Lesegenauigkeit und Leseverständnis erfasst (ebd.).

Eine einheitliche Operationalisierung wird allerdings nicht vorgeschrieben, so dass grundsätzlich unterschiedliche Verfahren zur Diagnose verwendet werden können (Scheerer-Neumann, 2018). Die ICD-10 geht davon aus, dass die Leistungen leseschwacher Kinder zwei Standardabweichungen (2.3%) unter dem Niveau liegen müssen, das aufgrund der Altersnorm und der allgemeinen Intelligenz erwartet wird (Mayer, 2016, S. 49). In vielen Tests wird allerdings bereits bei einer Standardabweichung (ca. 15%) von unterdurchschnittlichen Leistungen ausgegangen (Klicpera, Schabmann, Gasteiger-Klicpera & Schmidt, 2017, S. 136; Marx, 2007b, S. 81).

Gemäß Definition nach ICD-10 gehen mit Lesestörungen häufig auch Rechtschreibstörungen einher, so dass die Lesestörung nur in Kombination mit der Rechtschreibstörung beschrieben wird. Die Rechtschreibstörung hingegen wird auch als isolierte Störung definiert (Dilling et al., 2011). Die Frage, ob die beiden Formen der Beeinträchtigung überhaupt isoliert vorkommen,

⁵ Die ICD-10 wird derzeit überarbeitet und die Folgeversion ICD-11 wird 2022 in Kraft treten (Deutsches Institut für Medizinische Dokumentation und Information (DIMDI), 2019)

⁶ In der ICD und DSM wird meistens der Begriff „Störung“ verwendet, weshalb er hier, wie in den Originalen zitiert, verwendet wird.

wird in den Studien von Moll & Landerl (2011) und Winkes (2014) differenziert betrachtet. Moll & Landerl (2011, S. 22) weisen nach, dass ein beträchtlicher Anteil (40%) von Kindern mit Leseschwierigkeiten keine Probleme im Rechtschreiben aufweist. Aufgrund dieser Studie ist also davon auszugehen, dass es eine isolierte Lesestörung gibt, obschon diese bislang in der internationalen Fachliteratur als umstritten galt (ebd.). Die Gruppe der kombinierten Lese-Rechtschreibstörung ist mit 8% etwa gleich groß wie die jeweils isolierten Störungen „Lesen“ (6%) und „Rechtschreiben“ (7%). In der ICD-11 wird aufgrund neuer Erkenntnisse die Beschreibung einer isolierten Lesestörung erwartet (Bundesverband Legasthenie und Dyskalkulie, 2018).

Die in der ICD-10 enthaltene „IQ-Diskrepanzdefinition“ hat in der Fachwelt zu starker Kritik geführt (siehe Scheerer-Neumann, 2018; Schneider, 2017), auf die hier nicht weiter eingegangen wird. Interessant ist aber, dass im DSM-5 auf die Diskrepanzdefinition verzichtet wird und der Fokus mehr auf die Alters-Diskrepanz gelegt wird (Klicpera et al., 2017; Mayer, 2016). Im DSM-5 wird zudem bereits in der aktuellen Version eine isolierte Leseschwierigkeit beschrieben (Klicpera et al., 2017). Mit Bezug auf die Lesestörung werden hier besondere Schwierigkeiten in den drei Symptombereichen *Lesegeschwindigkeit*, *Lesegenauigkeit* und *Leseverständnis* sowie Probleme bei der Worterkennung und Graphem-Phonem-Zuordnung genannt (Schulte-Körne, 2014b). Die Beeinträchtigung muss gemäß DSM-5 länger als sechs Monate anhalten, auch bei Durchführung von lesebezogenen Interventionen. Zur Diagnose wird neben standardisierten Tests auch eine klinische qualitative Diagnose mit einbezogen (z.B. Familienanamnese, Schulbericht, Beurteilungsskalen). Zur Überprüfung quantitativer Kriterien wird empfohlen, standardisierte Tests einzusetzen, die Ergebnisse mit zu erwartenden Klassen-/Altersnormen zu vergleichen und eine Diskrepanz von ein bis zweieinhalb Standardabweichungen zu verwenden. Eine geringe Diskrepanz (eine Standardabweichung) kann dann angewendet werden, wenn klinische qualitative Kriterien die Diagnose stützen (ebd.). Zusammenfassend stellt Mayer (2016, S. 47) mit Bezug auf die Leitlinien der Deutschen Gesellschaft für Kinder- und Jugendpsychiatrie (2015) fest, dass zwischen Kindern und Jugendlichen, bei denen die Diagnose einer Lese-Rechtschreibstörung aufgrund einer Alters- oder Klassennormdifferenz gestellt wurde (DSM-5), und Kindern, bei denen die Diagnose auf dem IQ-Diskrepanzkriterium basierte (ICD-10), keine eindeutigen Unterschiede festgestellt werden konnten.

Die Prävalenz „beim Erwerb der Schriftlichkeit wird in Ländern mit alphabetischen Schriftsystemen auf 5-10% geschätzt“ (Scheerer-Neumann, 1997, S. 297). Etwas genauere Angaben finden sich bei Galuschka und Schulte-Körne (2016, S. 279), die feststellen, dass gemäß ICD-10 eine Prävalenz der *isolierten* Lesestörung um 6% und der *kombinierten Lese-Rechtschreib-*

störung von 8% vorliegt. Schließlich wird mit Bezug auf den Bereich des *Lesens* auf Grundlage aktueller Studien von einer Prävalenz von 3 - 8% ausgegangen. Damit gehört die Leseschwierigkeit zu einer der häufigsten Entwicklungsstörungen (Deutsche Gesellschaft für Kinder- und Jugendpsychiatrie Psychosomatik und Psychotherapie, 2015; Galuschka & Schulte-Körne, 2015).

Die beiden internationalen Klassifikationssysteme beschreiben das Phänomen der Leseschwierigkeit eher allgemein, indem die betreffenden Bereiche sowie Möglichkeiten der Erfassung (z.B. standardisierte Tests) genannt werden. Übereinstimmung besteht darin, dass sich die Beeinträchtigung der Lesekompetenz auf das sinnverstehende Lesen niederschlägt und dass damit die zentrale *Funktion* des Lesens eingeschränkt ist. Für die pädagogische Praxis dürften präzisere Beschreibungen der Symptomatik jedoch gewinnbringender sein, weshalb im Folgenden eine genauere Symptombeschreibung erfolgt, die zudem auf empirischen Untersuchungen basiert.

2.3 Komponenten der Lesekompetenz

2.3.1 Theoretische Verortung der Lesekompetenz

Zur Frage, was Lesekompetenz beinhaltet, werden in der Lesekompetenzforschung verschiedene Theorien und Ansätze diskutiert. Dabei liegen im deutschsprachigen Raum mehrere, sich oftmals überschneidende Modellierungen vor (Husfeld & Lindauer, 2009). Je nach Theorie und Forschungsziel der verschiedenen Ansätze werden in Bezug auf die Definition von Lesekompetenz unterschiedliche Akzente gesetzt, und die Frage nach der Definition von Lesekompetenz und dem Grundbegriff des Lesens wird zu unterschiedlichen Antworten führen (Hurrelmann, 2011).

Die im Folgenden skizzierten Modelle sollen eine Vorstellung darüber vermitteln, unter welcher Zielperspektive „Lesen“ betrachtet werden kann und auf welcher unterschiedlichen Ebenen das Konstrukt beschrieben wird. Dabei besteht hinsichtlich der erwähnten Modelle kein Anspruch auf Vollständigkeit.

Prominent werden das *kognitionstheoretische Modell* der PISA-Studie und das *kulturwissenschaftliche Modell* im Kontext der Lesesozialisationsforschung diskutiert (Hurrelmann, 2011). Im *kognitionstheoretischen* Modell, das in der psycholinguistischen Forschung beheimatet ist, spielen alltagsfunktionale Aspekte eine bedeutende Rolle, d.h. die kognitiven Dimensionen des Textverstehens, die sich vor allem auf Sach- und Informationstexte konzentrieren (Bertschi-Kaufmann & Kappeler, 2010; Hurrelmann, 2011). Lesekompetenz setzt sich in diesem Modell aus drei zentralen Dimensionen zusammen: 1) Informationen ermitteln, 2) textbezogenes

Interpretieren und 3) Reflektieren und Bewerten (Hurrelmann, 2011, S. 24). Als exemplarisches Modell, das in der Tradition der Psycholinguistik anzusiedeln ist, ist das *informationsverarbeitungstheoretische Modell* von Wember (1999) zu nennen. Neben der Zielperspektive (Informationsentnahme aus einem Text) werden bei diesem Ansatz vor allem die beim Lesen ablaufenden Verarbeitungsprozesse betont. Lesen bezeichnet somit „ein interaktives Deuten von schnell wahrgenommenen Textteilen auf der Grundlage des Vorwissens der Lesenden“ (ebd., S. 18). Demgegenüber steht der Lesebegriff der Lesesozialisationsforschung, der „die aktive und konstruktive Leistung des Lesers“ akzentuiert (Hurrelmann, 2011, S. 24). In diesem Modell werden nicht nur die kognitiven Aspekte, sondern auch die motivational-emotionalen und kommunikativ-interaktiven Fähigkeiten einbezogen (Bertschi-Kaufmann & Kappeler, 2010). Lesekompetenz wird hier definiert als „Fähigkeit zum Textverstehen im Horizont einer kulturellen Praxis, zu der es gehört, dass sich (1) *kognitives Textverständnis*, (2) *Motivation und emotionale Beteiligung*, (3) *Reflexion und Anschlusskommunikation* (mit anderen Lesern) ergänzen und durchdringen“ (Hurrelmann, 2011, S. 24). Das Modell bezieht kognitive hierarchieniedrige Prozesse des Lesens mit ein, erweitert diese aber um hierarchiehöhere Prozesse des Lesens.

Ein Modell, das aufgrund seiner didaktischen Verwendbarkeit Verbreitung gefunden hat, ist das *Mehrebenenmodell des Lesens* (Abbildung 1) von Rosebrock und Nix (2012), in dem Lesen aus einer didaktischen Perspektive betrachtet wird. Im Modell von Rosebrock und Nix wird ein umfassender Begriff von Lesekompetenz entworfen, der verschiedene Ebenen berücksichtigt und damit Aspekte der vorgängig genannten Modelle mit einbezieht. Insbesondere werden hier Dimensionen aus der Lesesozialisationsforschung und dem kognitionstheoretischen Modell verbunden (Nix, 2010). Da „Lesen“ bei diesem Ansatz aus einer didaktischen Perspektive betrachtet wird und das Modell „eine handlungsorientierte Beurteilung und Anleitung lesebezogener Unterrichtsprozesse im Rahmen des Deutschunterrichts ermöglicht“ (ebd. S. 144), ist es für die vorliegende Arbeit im Hinblick auf die Zielgruppe der Lehrpersonen besonders gewinnbringend.

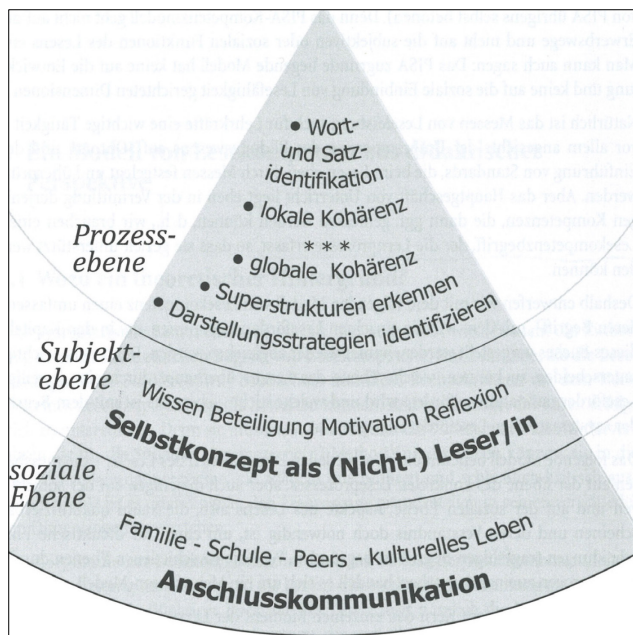


Abbildung 1. Mehrebenenmodell des Lesens (Rosebrock & Nix, 2012, S. 11).

Lesen setzt sich im Mehrebenenmodell von Rosebrock & Nix (2012) aus drei Dimensionen zusammen:

- 1) Einerseits ist die messbare *Prozessebene* angesprochen, welche die kognitiven Tätigkeiten bezeichnet, die während des Lesevorganges ablaufen (Rosebrock & Nix, 2012). Dieser Bereich betrifft die kognitiven Prozesse, die beim Wort-, Satz- und Textlesen ablaufen und umschreibt den Lesebegriff, wie er im kognitionstheoretischen Modell von PISA verwendet wurde (ebd.).
- 2) Im Weiteren sind Lesende immer auch als *Subjekt* am Leseprozess beteiligt und bringen ihr Vorwissen, ihre Motivation und ihr Selbstkonzept als Lesende in die Situation mit ein.
- 3) Im äußersten Kreis ist die *soziale Ebene* angesprochen, die Ebene, welche die Genese der Lesefähigkeit und Lesebereitschaft mitbestimmt (ebd.). Lesen setzt sich hier aus mehreren interagierenden Teilfähigkeiten zusammen, wobei diese in hierarchieniedrige und hierarchiehöhere Prozesse unterschieden werden (Lenhard & Artelt, 2009).

Im gezeigten Modell sind auf Prozessebene hierarchieniedrige Prozesse durch die Wortidentifikation und das Bilden lokaler Kohärenzen bezeichnet. Diese laufen weitgehend automatisiert ab. Die hierarchiehöheren Prozesse bewegen sich auf der Stufe globaler Kohärenzbildungen und erfordern während des Lesens eine bewusste gedankliche Anstrengung (z.B. Lesestrategien) (Rosebrock & Nix, 2012; Schneider et al., 2013).

Das Modell dient für die folgenden Ausführungen als Rahmen, da es verschiedene theoretische Ansätze verbindet und der Lesebegriff umfassend beschrieben ist. Der Fokus wird in der

vorliegenden Arbeit vor allem auf die hierarchieniedrige Prozessebene gelegt, weitere Ebenen sind aber teilweise mit einbezogen. Vor diesem Hintergrund werden im Folgenden zentrale Komponenten des Lesens besprochen und deren Zusammenhänge geklärt.

2.3.2 Hierarchieniedrige Prozesse: Basale Lesefertigkeiten und Leseflüssigkeit

Unter hierarchieniedrigen Prozessen werden basale Lesefertigkeiten (Buchstaben- und Wortebene) und die Lesegeläufigkeit bzw. Leseflüssigkeit (Wort- und Satzebene) eingeordnet (Kruse, Rickli, Riss & Sommer, 2010). Weiter wird das Erstellen kleinräumiger Textzusammenhänge (lokale Kohärenz) zu den hierarchieniedrigen Leseprozessen gezählt. Hierarchiehöhere Prozesse setzen hingegen an der Textbasis an und transformieren sie, damit ein umfassendes Textverstehen als Ganzes (globale Kohärenz) ermöglicht wird (Herné & Löffler, 2014). Dabei erleichtern gut ausgebildete basale Lesefertigkeiten die Entwicklung hierarchiehöherer Prozesse und damit das Textverständnis (Rosebrock & Nix, 2012).

Im Folgenden werden zunächst Prozessmodelle basaler Lesefertigkeiten diskutiert, und im Anschluss daran wird das Konstrukt der Leseflüssigkeit umrissen.

Prozessmodelle zu basalen Lesefertigkeiten (Wortebene)

Auf der Ebene basaler technischer Lesefertigkeiten sind zunächst Kompetenzen zur Identifikation von Wörtern angesprochen (Scheerer-Neumann & Hoffmann, 2005). Dies setzt voraus, dass Kinder Buchstaben(gruppen) und deren lautliche Zuordnung kennen, was fachsprachlich auch „Graphem-Phonem-Korrespondenz“ (GPK) genannt wird. Phoneme stehen dabei für die „kleinste bedeutungsunterscheidende Einheit der Sprache, Laute (synonym „Phone“) ihrer hörbaren Realisierung (Marx, 2007b, S. 23). Grapheme bezeichnen den mit einem Phonem korrespondierenden Buchstaben oder die korrespondierende Buchstabengruppe (ebd.).

Es existieren verschiedene Modellannahmen darüber, wie der Prozess des Wortlesens abläuft, wobei „Top-Down Modelle“ den Aspekt der Kontextnutzung betonen und von einem inneren lexikalischen Wortspeicher ausgehen, während in „Bottom-up-Modellen“ der systematische Aufbau von der Buchstaben- zur Wortebene angesprochen ist (Scheerer-Neumann, 1997; Schröder-Lenzen, 2013). Interaktive Modelle gehen von einer gegenseitigen Beeinflussung von Satzkontext und Worterkennung aus (Scheerer-Neumann, 1997), d.h., dass hierarchiehöhere und hierarchieniedrige Prozesse parallel oder in zeitlicher Überlappung durchlaufen werden (Richter & Christmann, 2002). Nach heutigem Erkenntnisstand kann davon ausgegangen werden, dass der *direkte lexikalische* Weg (top-down), bei dem Sinnerwartungen beim Lesen für die Entschlüsselung von Wortbedeutungen zentral sind und der *indirekte-phonologische* Weg (bottom-up), bei dem die Graphem-Phonem-Zuordnung sukzessive zur Entschlüsselung

von Wörtern genutzt wird, interagieren (Schröder-Lenzen, 2013). Solche interaktiven informationstheoretischen Modelle des Worterkennens werden auch als *Zwei-Wege-Modelle* bezeichnet und basieren auf der Vorstellung, dass Lesende „entweder direkt einen Kontakt mit dem lexikalischen Eintrag des Wortes herstellen, und zwar über den orthographischen Kode des Schriftbildes, oder aber den indirekten Weg über die Phonemfolge gehen“ (Klicpera et al., 2013, S. 51). Ein Modell, das für das Lesen als gut abgesichert gilt (Klicpera et al., 2013, 51; Winkes, 2014) und für die Praxis tragfähig ist (Günther, 2007), ist das „Dual Route Cascaded Model“ von Coltheart (2005) bzw. die Modelle, die auf diesen Grundannahmen beruhen. Der Rückbezug auf Zwei-Wege-Modelle ist insbesondere aufgrund seines Nutzens für das Verständnis von Schwierigkeiten lese- bzw. schreibschwacher Kinder sinnvoll und hat deshalb einen stärkeren Einfluss als alternative Modelle wie z.B. konnektionistische Netzwerkmodelle (Klicpera et al., 2013). Es gibt mittlerweile verschiedene Versionen dieses Zwei-Wege-Modells, die sich jeweils lediglich in Details unterscheiden (Winkes, 2014). Eine Darstellung eines Zwei-Wege-Modells, das auf den Leseprozess fokussiert, findet sich bei Scheerer-Neumann (1998) und soll exemplarisch für andere Darstellungen stehen (Abbildung 2).

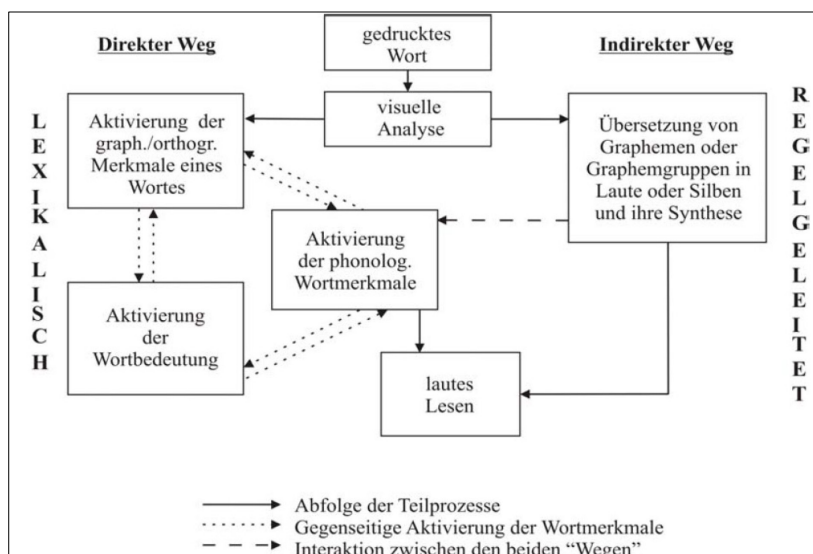


Abbildung 2. Zwei-Wege-Modell des Wortlesens (Scheerer-Neumann, 1998, S. 15)

Im ersten Schritt werden visuelle Eigenschaften eines Wortes analysiert, danach divergieren die Wege je nachdem, ob das Wort bekannt ist oder nicht (Klicpera et al., 2013). Beim *direkten Leseweg* wird das Wort als Ganzes in einem inneren orthographischen (mental) Lexikon abgespeichert und in der Lesesituation als Ganzes abgerufen (Günther, 2007). Dieser Weg wird in Entwicklungsmodellen (Kap. 2.6) der *orthographischen Strategie* zugeordnet. Daneben muss bei beim Lesen richtiger Wörter bei einer bestimmten Buchstabenfolge bzw. einem Wort das semantische System aktiviert werden, das die Bedeutung und den Inhalt des gelesenen

Wortes vermittelt (Günther, 2007; Klicpera et al., 2013). Der *indirekte Weg* erfolgt demgegenüber unter Anwendung der Graphem-Phonem-Korrespondenz. In Entwicklungsmodellen wird dies als *alphabetische Strategie* bezeichnet (Kap. 2.6). Hier müssen Grapheme in die entsprechenden Phoneme umgewandelt werden, um schließlich koartikulatorisch zu einem Wort synthetisiert zu werden; dabei führt der Weg zunächst am semantischen Weg vorbei (Mayer, 2016). Erst in dem Moment, in dem das phonologisch rekodierte Wort semantisch verknüpft und damit lexikalisch dekodiert und schließlich artikuliert werden kann, ist der Lesevorgang abgeschlossen (Wember, 1999). Dieser Prozess verläuft bei Leseanfängern und Leseanfängerinnen langsam und mühsam und beansprucht das Arbeitsgedächtnis und die kognitiven Ressourcen, die dann für das Leseverständnis oft nicht mehr zur Verfügung stehen (Mayer, 2016). Das Problem stellt sich darin, dass für die Informationsverarbeitung (Visuelle Perzeption – Phonematische Rekodierung – Lexikalische Dekodierung – Syntaktische und semantische Interpretation) das Kurzzeitgedächtnis als Arbeitsspeicher und zudem die Aufmerksamkeit als Instanz für die Regulation des Lesevorgangs benötigt werden (Wember, 1999). Dies kann für Kinder mit Lese-schwierigkeiten eine entscheidende Hürde sein (ebd.).

Entwicklungsmodelle aus der Informationsverarbeitungstheorie spielen heute unbestritten eine herausragende Rolle (Kirschhock, 2004). Die Grundannahme des Erlesens von Wörtern nach dem Zwei-Wege-Modell ist zwar stark vereinfacht und wurde in den letzten Jahren auch kritisiert (Klicpera et al., 2013; Mayer, 2016). Das Modell stellt aber dennoch grundlegende Punkte basaler Leseprozesse dar (Marx, 2007b). Die Kritik betrifft insbesondere die Annahme zweier weitgehend unabhängiger Verarbeitungswege für bekannte Wörter (direkter Weg) auf der einen Seite und Pseudowörter bzw. Ausnahmewörter (indirekter Weg) auf der anderen Seite (Mayer, 2016). Zudem vernachlässigt dieser Ansatz den Kontext, der beim Lesen eine nicht zu unterschätzende Rolle spielt (Günther, 2007).

Gegenentwürfe zu diesem Modell stellen sogenannte *konnektionistische Modelle* dar (Kirschhock, 2004). Diese Theorie des einfachen Zugangswegs, die prototypisch von Seidenberg und McClelland (1989) entwickelt wurde, geht davon aus, dass für das Erlesen von Wörtern nur ein einziges Verarbeitungssystem erforderlich ist (Klicpera et al., 2013; Mayer, 2016). Gemäß dieser Theorie bauen Lesende durch Erfahrung mit Schriftsprache ein Netzwerk auf, in dem verschiedene Informationen über Wörter (Orthographie, Phonologie, Wortbedeutung) abgebildet sind, auf die sie unbewusst und rasch zugreifen können (Klicpera et al., 2013). Der Korpus an Regelmäßigkeiten aber auch Unregelmäßigkeiten und Ausnahmen der Schriftsprache wird im Sinne „statistischen Lernens“ (Winkes, 2014, S. 22) verwendet, wobei auf die

Verknüpfung von orthographischen und phonologischen Einheiten fokussiert wird (Mayer, 2016).

Damit ergibt sich ein weiterer Blick auf den basalen Leseerwerb, wobei in allen skizzierten Modellen *die Automatisierung* von Prozessen zentral ist. Die beschriebenen Modelle beziehen sich zunächst auf das Wortlesen. Die nächste Ebene des basalen Leseerwerbs fokussiert auf die Satz- und Textebene, wobei der Begriff der Leseflüssigkeit zentral ist. Die Leseflüssigkeit nimmt eine „Brückenfunktion“ zwischen Wortlesen und Textverständnis (hierarchiehoher Prozess) ein und bezeichnet den „Flaschenhals“ zum verstehenden Lesen (Hartmann & Niedermann, 2008, S. 5; Scheerer-Neumann, 2015, S. 73).

Leseflüssigkeit (Satz- und Textebene)

Zu den hierarchieniedrigen Prozessen gehören neben den eben gezeigten Lesefertigkeiten automatisierte Abläufe beim Satzlesen (Kruse et al., 2010). Das flüssige Lesen hängt dabei eng mit der Entwicklung des direkten lexikalischen Lesewegs zusammen, wie er im *Dual Route Model* besprochen wurde (Rosebrock, Nix, Rieckmann & Gold, 2011). Dem Bereich der Leseflüssigkeit wird in der Leseforschung besondere Beachtung geschenkt, nicht zuletzt, weil Ergebnisse internationaler und deutschsprachiger Studien zeigen, dass viele Kinder und Jugendliche über mangelnd ausgebildete Lesefertigkeiten auf Wort- und Satzebene verfügen und Texte nicht flüssig lesen können (ebd.).

In Bezug auf die Bedeutung der Leseflüssigkeit sind im deutschsprachigen Raum insbesondere die Längsschnitt-Studien von Landerl und Wimmer (2008) sowie von Klicpera und Gasteiger-Klicpera (1993) aufschlussreich. Im englischsprachigen Raum ist u.a. die Metaanalyse des *National Reading Panel* (National Reading Panel, 2000) bedeutend. Mit Bezug auf Studien aus dem englischsprachigen Raum fasst Rasinski (2017) zusammen:

These studies and others (e.g., Rasinski & Padak, 1998) have suggested that difficulties in the foundational competencies (word recognition and fluency) are a major contributor to reading difficulties early on and that if students do not develop early mastery of these foundational reading competencies, it is likely that these concerns will continue into the later grades and will have a profound, adverse effect on students' comprehension and overall reading achievement. (S. 521)

Leseflüssigkeit – auch als „*reading fluency*“ bezeichnet – meint das mühelose und routinierte Lesen von Texten, ohne dass Lesende merken, dass sie lesen (Rosebrock et al., 2011). Dabei umfasst diese Kompetenz verschiedene zusammenhängende Dimensionen: 1) Genauigkeit des Dekodierens, 2) Automatisierung des Dekodierens, 3) Lesegeschwindigkeit und 4) pro-

sodisches, sinngestaltendes Lesen (ebd., S.15ff.). Die Genauigkeit und der Automatisierungsgrad des Dekodierens betreffen die Wortebene; die Lesegeschwindigkeit und das betonte Lesen gehen hingegen über die Wortgrenze hinaus (Gold, 2009). Die folgende Kategorisierung geht auf die theoretische Modellierung des Begriffs „*Fluency*“ aus dem angloamerikanischen Raum zurück (Rosebrock & Nix, 2006, S. 3).

1) *Genauigkeit des Dekodierens*: Die Lesegenauigkeit bezieht sich auf das korrekte Re- und Dekodieren von Wörtern und ist für die Sinnentnahme beim Lesen von Texten entscheidend. Dies wird insbesondere dann sichtbar, wenn ein kritischer Wert an Dekodiergenauigkeit (90%) unterschritten wird, wie Rosebrock et al. (2011, S. 16) mit Bezug auf die Studie von Rasinski (2003) festhalten. Für ein unabhängiges Lesen wird meist der Wert von mindestens 95% korrekt gelesener Wörter angenommen, dagegen werden Wertebereiche zwischen 90%-95% als „Instruktionsniveau“ bezeichnet (Hartmann & Niedermann, 2008, S.17f.; Rosebrock et al., 2011, S. 61). Texte, die auf Instruktionsniveau gelesen werden, bieten eine intellektuelle Herausforderung, können aber mit geeigneter Unterstützung bewältigt werden (Rosebrock et al., 2011). Bei der Entwicklung der Lesegenauigkeit zeigt sich, dass sie bei durchschnittlich lesenden Kindern bereits in der zweiten Klasse bei 95% liegt und dass Ende der dritten Klasse ein gewisses Plateau erreicht wird, während sich die Lesegeschwindigkeit über die gesamte Schulzeit weiterentwickelt (Marx, 2007b; Richter & Müller, 2017). Lesegenauigkeit und Lesegeschwindigkeit bedingen einander. In der Regel scheint der Fokus aber zunächst auf der Genauigkeit zu liegen, d.h. die meisten Kinder entwickeln bis Ende der ersten Klasse eine hohe Sicherheit in der Beherrschung von Buchstaben und dem Lesen von kürzeren und häufigen Wörtern (Klicpera et al., 2013).

2) *Automatisierung*: Automatisierung des Dekodierens bedeutet, dass Kinder Wörter zunehmend auf direktem lexikalischen Weg lesen und dass die Worterkennung in immer größeren sublexikalischen Einheiten (Silben, Morpheme, Signalgruppen) erfolgt (Rosebrock, Rieckmann, Nix & Gold, 2010; Wember, 1999). Der Zuwachs an Geschwindigkeit hängt damit zusammen, dass die Informationsverarbeitung durch die „Bündelung“ von Einheiten beschleunigt wird. Klicpera et al. (2013) sprechen in diesem Zusammenhang auch von „partiell lexikalischem Lesen“ (S. 34). Im Rahmen der später in dieser Arbeit dargestellten Stufenmodelle wird hier der Übergang von der alphabetischen hin zur orthographischen Lesestrategie markiert. Der Aufbau eines orthographischen Lexikons (Sichtwortschatz) ist für die direkte Worterkennung grundlegend (Hartmann & Niedermann, 2008). Übereinstimmend wird der Prozess der Automatisierung des Wortlesens als zentrales Moment der Leseentwicklung bezeichnet, da ein Mangel an Automatisierung zur Bindung kognitiver Kapazitäten führt, die

dann für nachfolgende hierarchiehöhere Prozesse nicht mehr zur Verfügung stehen (Hartmann & Niedermann, 2008; Rosebrock & Nix, 2012; Wember, 2015). Bei leseschwachen Kindern zeigt sich oft genau hier ein grundlegendes Defizit, da ein automatisierter Leseprozess durch ein nur unzureichend ausgebildetes inneres Lexikon nachhaltig verzögert wird (Schründer-Lenzen, 2013).

3) *Lesegeschwindigkeit*: Die Lesegeschwindigkeit ist gewissermaßen die logische Konsequenz aus einer genaueren Worterfassung und einem höheren Automatisierungsgrad (Rosebrock et al., 2011) und bezieht sich auf die Satz- und Textebene (Rosebrock & Nix, 2012). Operationalisiert wird sie meist über die Anzahl der korrekt gelesenen Wörter in einem bestimmten Zeitintervall (Müller, Krizan, Hecht, Richter & Ennemoser, 2013). Damit von *flüssigem* Lesen gesprochen werden kann, wird bei Erwachsenen in der Regel von einem „Normalmodus“ von 250-300 Wörtern pro Minute ausgegangen (Rosebrock et al., 2011, S. 18). Allerdings erfolgt der Übergang zum sinnentnehmenden flüssigen Lesen schon bei etwa 100 Wörtern pro Minute (Rosebrock & Nix, 2012, S. 32). Diese „Betriebsgeschwindigkeit“ wird von durchschnittlich Lesenden etwa Ende der dritten Klasse erreicht (Klicpera & Gasteiger-Klicpera, 1993; Marx, 2007b), hängt aber in hohem Masse auch von der Textschwierigkeit ab (Kap. 2.7). Gemäß Wember (1999) lesen Kinder, die weniger als 25 Wörter pro Minute lesen und in dieser Zeit mehr als acht Wörter fehlerhaft lesen, auf dem „Überforderungsniveau“ (S. 36). Für diese Kinder muss das Textmaterial entsprechend angepasst werden. Ab einem Richtwert von 100 Wörtern pro Minute und einer Genauigkeit von 95% korrekt gelesener Wörter kann ein „unterhaltender Fließtext auf Unabhängigkeitsniveau verstehend gelesen werden“ (Rosebrock et al., 2011, S. 62).

Die Bedeutung des schnellen Lesens lässt sich zweifach begründen: Zum einen entstehen beim (zu) langsamen Lesen Verstehensprobleme, da Informationseinheiten nicht zur selben Zeit im Kurzzeitgedächtnis gespeichert werden können (Rosebrock et al., 2011; Wember, 1999). Zum anderen „erschwert eine langsame Lesegeschwindigkeit die Effizienz der Selbstüberwachungsprozesse beim Lesen“ (Rosebrock et al., 2011, S. 18). Daraus sollte jedoch nicht der Schluss gezogen werden, dass je schneller desto besser gelesen wird. Vielmehr ist ein dem Textkontext angemessenes Lesetempo anzustreben, das die Informationsentnahme mühelos ermöglicht (Rosebrock & Nix, 2012). Für das Erreichen flüssigen Lesens, bei dem die Aufmerksamkeit für das Verstehen von Texten aktiviert werden kann, kommt es nicht nur darauf an, das Lesetempo durch Automatisierung des Dekodierprozesses zu steigern, sondern auch darauf, dass Lesende ihr Tempo der jeweiligen Gegebenheit anpassen und das Lesen *flexibilisieren* (Holle, 2006). Verschiedene Studien (u.a. Klicpera, Ehgartner, Gasteiger-Klicpera &

Schabmann, 1993) zeigen übereinstimmend, dass Defizite in der Leseflüssigkeit über die Zeit sehr stabil bleiben, was sich darin äußert, dass die Gruppe der sehr schwachen Lesenden auch nach achtjähriger Beschulung langsamer lesen als die Gruppe der guten Leserinnen und Leser in der zweiten Klassenstufe (Schneider, 2009). Zu einem ähnlichen Ergebnis kommen Landerl und Wimmer (2008), die ebenfalls eine hohe Stabilität von Defiziten in der Leseflüssigkeit feststellen.

4) *Prosodie*: Die vierte Ebene der Leseflüssigkeit betrifft die Fähigkeit, Texte gestaltend vorzulesen (Husfeld & Lindauer, 2009). Die Prosodie – im Sinne eines natürlichen Ausdrucks beim lauten Lesen – impliziert Merkmale wie „Intonation, Betonung, Rhythmus auf der Phrasen-, Satz- und Textebene“ (Hartmann & Niedermann, 2008, S. 7). Eine qualitative Erfassung verschiedener Stufen prosodischer Kompetenzen wurde durch die *Oral Fluency Scale* von Pinnell (1995) angeregt und im Rahmen neuerer Studien weiterentwickelt (Nix, 2011; Sappok & Fay, 2018). Bei Rosebrock et al. (2011, S. 86) werden in Anlehnung an die Skala von Pinnell vier Niveaus der Prosodie operationalisiert: Die beiden untersten Stufen bezeichnen Leseprozesse, die noch weit weg sind von einer flüssigen Lektüre und bei denen ohne angemessene Segmentierung Wort-für-Wort oder in Zweiwortgruppierungen gelesen wird. Kinder, die auf Stufe drei und vier lesen, können als flüssig lesende Kinder charakterisiert werden, bei denen eine expressive Gestaltung einzelner bzw. mehrerer Textabschnitte vorliegt und deren Leseablauf durch größere Wortgruppierungen der Syntax des Textes überwiegend oder vollständig entspricht (ebd.).

Prosodisches Lesen ist insbesondere für das Textverstehen bedeutsam, da durch sinnvolle Wortgruppierungen, Pausengestaltung usw. Sätze schon während des Lesens in kleinere kohärente Sinneinheiten zerlegt werden (Rosebrock & Nix, 2006). Allerdings stellt sich hier die Frage der Kausalität, da das sinnvoll betonte Lesen auch eine Folge des besseren Leseverständnisses sein könnte und damit auch reziprok verlaufende Wirkmechanismen angenommen werden können (Gold, 2009; Rosebrock & Nix, 2006). Die diesbezügliche Forschungslage ist nicht eindeutig (Hartmann & Niedermann, 2008). Selbst Zusammenhänge (Korrelationen) zwischen Prosodie und Leseverstehen sind nicht durchwegs zu finden. Während Rasinski (2017) aufgrund von Studien von einer hohen Korrelation zwischen prosodischen Merkmalen und dem Leseverstehen („silent reading comprehension“) ausgeht, kann der Zusammenhang zwischen prosodischem Ausdruck und Leistungen im Satz- und Wortverständnis (stilles Lesen) in der Frankfurter-Studie (Gold, 2009) nicht bestätigt werden. Gemäß Ergebnissen der Frankfurter-Studie lässt sich aber zumindest ein korrelativer Zusammenhang zwischen den verschiedenen Komponenten der Leseflüssigkeit feststellen: „Ein geringerer Automati-

sierungsgrad des Dekodierens geht also mit vielen sinnentstellenden Lesefehlern einher, und wer vergleichsweise schneller lesen kann, liest meist auch mit angemessener Betonung“ (Gold, 2009, S. 155). Der Blick auf Kinder mit Leseschwierigkeiten zeigt, dass diese auch bezüglich mangelnder prosodischer Kompetenzen beim lauten Vorlesen auffallen (Mayer, 2016; Rasinski, 2017). Die Erklärung hierfür liegt in der erwähnten engen Verbindung der verschiedenen Dimensionen der Leseflüssigkeit. Wenn die Anforderungen für die Worterkennung (Genauigkeit und Automatisierung) hoch sind, bleiben meist weniger kognitive Kapazitäten für die bedeutungsvolle Textrepräsentation und Interpretation des Gelesenen übrig (Hartmann & Niedermann, 2008).

2.3.3 Hierarchiehohe Prozesse: Sinnverstehendes Lesen

Die hierarchiehöheren Prozesse betreffen das zielorientierte, reflektierende Lesen sowie die bewusste Planung und Überwachung des Lesevorgangs (Kruse et al., 2010). Auf dieser Ebene steht das *Leseverständnis* im Zentrum, sowie alle Prozesse und beteiligten Komponenten, die für das textverstehende, sinnentnehmende Lesen – als Ziel des Lesens schlechthin – beteiligt sind. Die hierarchiehöheren Prozesse beziehen sich auf die Textebene und ermöglichen ein globales Verständnis des Textganzen (globale Kohärenz), sie umfassen zudem Komponenten, die das Verstehen und Behalten von Texten erleichtern (Rosebrock et al., 2011). Zunächst soll ein Einblick in das Konstrukt des *Leseverständnisses* vermittelt und als ein Modell von Leseverständnis der „Simple View of Reading“-Ansatz umschrieben werden. Nachfolgend werden zentrale Komponenten, die im Zusammenhang mit dem Leseverständnis stehen, besprochen. Es sind dies die Aspekte *lokale und globale Inferenzbildung*, *Wortschatz/Weltwissen*, sowie *Lesestrategien* (Scheerer-Neumann, 2015). Zudem wird der Einfluss der *Lesemotivation* kurz skizziert.

Das Konstrukt des Leseverständnisses und der Simple View of Reading-Ansatz

Das Leseverständnis (Ergebnis) bzw. das Leseverstehen (Prozess) bezieht sich auf die Satz- und Textebene (Scheerer-Neumann, 2015). Das „Verstehen“ bildet dabei die Essenz des Lesens, das heißt: „Kompetente LeserInnen konstruieren laufend und effizient Bedeutung, gewinnen zielgerichtet Informationen und können durch Lesen ihr Wissen erweitern“ (Hartmann & Niedermann, 2008, S. 7). Auf dieser allgemeinen Ebene lassen sich zahlreiche vergleichbare Umschreibungen des Konstrukts finden (z.B. Klicpera et al., 2013; Marx, 2007b). Anders stellt sich die Situation in Bezug auf die beteiligten Komponenten des Leseverständnisses dar bzw. hinsichtlich des Fokus, der in verschiedenen Modellen gesetzt wird (Lenhard, 2019). Der *Simple View of Reading-Ansatz* (Gough & Tunmer, 1986) ist ein bekanntes Modell, das neben

Prozessen auf der Ebene der Worterkennung auch allgemeine Fähigkeiten zur Sprachrezeption berücksichtigt und einen erweiterten Blick auf beteiligte Teilprozesse beim Leseverstehen erlaubt (Lenhard, 2019). Aufgrund seiner herausragenden Bedeutung sollen dessen Grundzüge im Folgenden skizziert werden.

Dieser Ansatz wird im Zusammenhang mit dem Leseverständnis auch aktuell immer wieder diskutiert (Asadi & Ibrahim, 2018; Gold, 2009; Marx, 2007b; Mayer, 2016; Rosebrock et al., 2011) und bildet die Grundlage für zahlreiche Erweiterungen von Modellen zum Leseverständnis (u.a. Marx, 2007a; Walter, 2013). Seine empirische Evidenz wurde für verschiedene Sprachräume (siehe Forschungsüberblick Asadi & Ibrahim, 2018; Mayer, 2016; Walter, 2013) inklusive den deutschsprachigen Raum (Marx & Jungmann, 2000) bestätigt. Der Simple View of Reading-Ansatz gibt Einblick in zentrale Komponenten des Leseverständnisses (Mayer, 2016), auch wenn er aufgrund seiner starken Vereinfachung in der Kritik steht. Gemäß den Prämissen des Modells entwickelt sich das Leseverstehen in Abhängigkeit der *Worterkennung* (Rekodieren und Dekodieren) und des *Hörverstehens* (Marx & Jungmann, 2000). Das Leseverstehen lässt sich durch die Gleichung „Leseverständnis = Wortlesen x Hörverstehen“ operationalisieren (ebd., S. 115). Die Bedeutung der Worterkennung meint das automatisierte und schnelle Abrufen auf dem direkten oder indirekten Leseweg und impliziert damit wesentliche Bestandteile des Leseflüssigkeitskonstrukts (Gold, 2009; Walter, 2013). Die Worterkennung selbst speist sich aus weiteren Faktoren wie der Buchstabenkenntnis (Graphem-Phonem-Korrespondenz) und diese wiederum aus phonologischen Fertigkeiten. Die Worterkennung stellt dabei eine notwendige, aber nicht hinreichende Bedingung für das Leseverständnis dar (Mayer, 2016, 36). Das Hörverstehen, als zweiter Faktor der Gleichung meint „alle sprachlich-kognitiven Fähigkeiten (...) die zur korrekten und situationsangemessenen Interpretation von Sprechakten benötigt werden“ (Mayer, 2016, S. 37). Dazu gehören demzufolge u.a. Wortschatz, Weltwissen und grammatikalische Kompetenzen.

Das Modell geht weiter davon aus, dass ein multiplikativer Zusammenhang zwischen den zwei Faktoren besteht und sich auch die Beeinträchtigung nur in einem Faktor auf das Leseverstehen insgesamt auswirkt (ebd.). Walter (2013) fasst dies folgendermaßen zusammen:

Eine defizitäre Wortlesefähigkeit (sehr langsames, stockendes, fehlerhaftes Dekodieren) erschwert oder verhindert (weil Aufmerksamkeit bzw. Arbeitsgedächtniskapazität absorbierend) das Zusammenziehen von Einzelinformationen (z.B. auf der Basis von Wörtern) zu größeren Sinneinheiten. Dies wiederum erschwert oder verhindert z. B. das Aktivieren von Vorwissen (über das Lese-Thema), die Integration von neuen Informationen in vorhandene

Wissensstrukturen und/oder das Initiieren von metakognitiven Kontrollprozessen (z.B. Vornehmen von Korrekturen). (S. 15)

Die Hypothese des Modells, dass sich Lesefähigkeit (im Sinne von Leseverständnis) einfach als *Produkt* von Dekodierfähigkeit und Hörverstehen beschreiben lässt, wurde in neueren Studien auch angezweifelt, nicht aber, dass die beiden Komponenten insgesamt zur Varianzaufklärung beitragen (Schneider, 2017). Zur Entwicklung der beiden Komponenten lässt sich festhalten, dass der Einfluss der Worterkennung auf das Leseverständnis in den ersten Schuljahren größer ist als das Sprach- bzw. Hörverständnis, dass sich dieses Verhältnis aber im Laufe der Zeit verändert und lexikalische und grammatische Fähigkeiten zunehmend wichtiger werden (Mayer, 2016). Das Hörverständnis bildet mit zunehmend besseren Fertigkeiten im Worterkennen (zumindest für das Grundschulalter) die jeweils obere Leistungsgrenze des Leseverstehens (Marx & Jungmann, 2000; Walter, 2013). Mit Blick auf den Zusammenhang von Hör- und Leseverständnis schließlich lässt sich feststellen, dass das Hörverständnis zu Beginn der Schulzeit in der Regel besser ausgebildet ist als das Leseverständnis. Diese Relation wendet sich aber mit zunehmenden Lesekompetenzen in den höheren Klassen ins Gegenteil, was bei unauffällig lesenden Kindern mit verbesserten Überwachungs- und Lesestrategien einhergeht (Klicpera et al., 2013).

Lokale und globale Kohärenzbildung

Auf hierarchiehöherer Ebene der Lesekompetenz spielen Inferenzen eine bedeutende Rolle. Die Inferenzbildung bezeichnet die Fähigkeit, den semantischen Gehalt der Wörter auf Satzebene miteinander in Bezug zu setzen (lokale Kohärenz) und darüber hinaus auf Textebene den sinnhaften Gesamtzusammenhang (globale Kohärenz) herzustellen (Lenhard & Artelt, 2009; McElvany & Schneider, 2009). Das bedeutet, dass implizite, verborgene Informationen erkannt und Schlussfolgerungen gezogen werden müssen (Schründer-Lenzen, 2013). Im Gegensatz zur lokalen Kohärenzbildung, bei der gemäß Schneider (2017) „semantische Relationen zwischen aufeinanderfolgenden Sätzen analysiert werden“ (S. 24), geht es bei der globalen Kohärenzbildung um die Bildung von Makrostrukturen und die Analyse von Zusammenhängen größerer Textteile auf höherer Abstraktionsebene (ebd.). Das Ziel sinnverstehenden Lesens ist der selbständige und kompetente Umgang mit Texten (McElvany & Schneider, 2009) und setzt dort an, wo hierarchieniedrige Prozesse (Leseflüssigkeit) beherrscht werden und wo Lesende Argumentationslinien, Hauptgedankengänge und die Intention des Geschriebenen erfassen können (Nix, 2011).

Lokale Kohärenzbildung: Auf der Ebene lokaler Kohärenzbildung spielt neben syntaktischem und grammatikalischem Wissen auch das semantische Wissen (Vorwissen, Wortschatz) eine bedeutsame Rolle, bzw. spielen diese Ebenen zusammen (Lenhard & Artelt, 2009; Richter & Christmann, 2002; Schröder-Lenzen, 2013). In diesem Zusammenhang trifft man oft auf den Begriff des „parsing“. Damit wird der Sachverhalt beschrieben, dass bereits beim Lesen jedem Wort eine bestimmte Position in der Struktur des Satzes zugewiesen wird (Klicpera et al., 2013, S. 73). Führt die angenommene Satzstruktur in eine „Sackgasse“ und nicht zum erwarteten Ziel, muss sie revidiert werden (Richter & Christmann, 2002, S. 30). So wird der Satz „Alex ist leicht zu finden“ von Kindern im Alter von sieben bis acht Jahren oft dahingehend verstanden, dass Alex andere Kinder leicht findet. Dies deshalb, weil „Alex“ aufgrund seiner Position im Satz als Subjekt und nicht als Objekt wahrgenommen wird (Lenhard, Schneider, Lenhard & Lippard, 2006). Nach interaktionistischer Auffassung hingegen werden syntaktische und semantische Informationen *parallel* verarbeitet und zudem vom pragmatischen Kontext und dem Vorwissen bestimmt (Richter & Christmann, 2002; Schröder-Lenzen, 2013). Mit Bezug auf das Korrekturverhalten bei semantisch-syntaktischen Fehlern zeigt sich, dass leseschwachen Kindern diese Selbstkorrektur oft nicht gelingt, da sie im Vergleich zu gut lesenden Kindern ihren Lesefehler gar nicht bemerken und entsprechen auch nicht korrigieren (Schröder-Lenzen, 2013).

Globale Kohärenzbildung: Auf satzübergreifender Ebene spielen zur Bildung globaler Kohärenzen zum einen sogenannte Kohäsionsmittel eine wichtige Rolle (Vorverweise, Rückverweise, Wortwiederholungen, zeitliche Gliederung usw.), um zu einer stimmigen Textrepräsentation zu gelangen (Lenhard et al., 2006; Richter & Christmann, 2002). Zum anderen nehmen der Wortschatz und das thematische Vorwissen eine herausragende Funktion ein (Schröder-Lenzen, 2013). Es besteht Einigkeit darüber, dass ein gut ausgebildeter Wortschatz einer der wesentlichsten Einflussfaktoren auf das Leseverständnis ist (Juska-Bacher, Beckert, Stalder & Schneider, 2016; Klicpera et al., 2013; McElvany & Schneider, 2009; Scheerer-Neumann, 2015; Walter, 2013). Ein umfassender Wortschatz erlaubt den schnellen und sicheren Zugriff auf das mentale Lexikon, wodurch Arbeitsgedächtniskapazitäten frei werden für höhere kognitive Prozesse. Gleichzeitig ist der Wortschatz auch Indikator für vorhandenes Hintergrundwissen, welches das Textverständnis erleichtert bzw. erst ermöglicht (McElvany & Schneider, 2009).

Wortschatz/Vorwissen

Der Wortschatz und das damit assoziierte thematische Vorwissen hängen einerseits mit basalen Lesefähigkeiten (Dekodieren), aber auch mit weiterführenden Lesefähigkeiten (globale Kohärenzbildung) zusammen. Ein Überblick zu den komplexen Zusammenhängen von inhaltlichem Vorwissen, Wortschatz, Lesegeschwindigkeit, Inferenzen und Leseverstehensstrategien, welche die zentralen Prädiktorvariablen des Textverstehens bilden, findet sich u.a. bei Richter und Christmann (2002), Lenhard und Artelt (2009) und Klicpera et al. (2013).

Ein differenzierter Blick auf Ergebnisse zum Zusammenhang von Wortschatz und Leseverständnis verdeutlicht, dass sich der Einfluss über die Zeit verändert und im Verlaufe der Grundschule (erste bis dritte Klasse) zunimmt (Seigneuric & Ehrlich, 2005) bzw. erst später überhaupt direkt wirksam wird (Juska-Bacher et al., 2016). Juska-Bacher et al. (2016) belegen in ihrer Studie, dass sich der Wortschatz in der Anfangsphase des Leseerwerbs (erste und zweite Klasse) nur indirekt über die phonologische Bewusstheit vermittelt auf das Wortlesen auswirkt. Dieses Ergebnis schließt gemäß Juska-Bacher et al. (2016) an frühere Studien an, die darauf hinweisen, dass der Wortschatzumfang für das Dekodieren, nicht aber für das Rekodieren bedeutsam ist und vor allem in einer späteren Leselernphase, „wenn eine gewisse Automatisierung der basalen Lesefertigkeiten eingetreten ist“ (S. 34) zunehmend an Bedeutung gewinnt. Die Bedeutung des Wortschatzumfangs für das Leseverstehen kann für höhere Stufen (dritte Klasse und höher) als gesichert gelten (Gold, 2009; McElvany, Becker & Lüdtke, 2009).

Mit Blick auf Kinder mit Leseschwierigkeiten zeigt sich auch hier, dass diese einen geringeren Wortschatz aufweisen (Wimmer, 1993), was sich in der Folge wiederum auf das Leseverständnis niederschlägt. An verschiedenen Stellen wird darauf hingewiesen, dass häufiges Lesen mit einem umfangreicheren Wortschatz einhergeht (McElvany et al., 2009) und dass damit die zunehmenden Unterschiede bezüglich Wortschatz und inhaltlichem Vorwissen zwischen lese-schwachen und lesestarken Kindern auch eine Folge von Leseschwierigkeiten sein könnten (Rosebrock & Nix, 2012; Wimmer, 1993).

Kognitive und metakognitive Lesestrategien

Weitere Komponenten, die für das sinnverstehende Lesen bedeutsam sind, sind kognitive und metakognitive Fertigkeiten, die dabei helfen, das Lesen zu überwachen und zu steuern (Schneider, 2017). Unter dem Begriff der Lesestrategien werden „bewusst gewählte und selbst kontrollierte Vorgehensweisen und Verstehensoperationen zusammengefasst, die systematisch und gezielt zum Verständnisaufbau und Gebrauch von Texten eingesetzt werden“ (Kruse, 2006, S. 1). Rosebrock und Nix (2012) bezeichnen Lesestrategien auch als „mentale Werkzeuge“ (S.

59), mit denen Lesende hierarchiehohe Verstehensanforderungen von Texten während des Lesens unterstützen können und die auf das Ziel der Sinnentnahme, der Identifizierung von Hauptgedanken, das Textverständnis, aber auch die Reduktion von Inhalten oder die Reorganisation von Texten ausgerichtet sind (ebd.). Der Einfluss von Lesestrategien auf das Leseverständnis ist in Interventionsstudien und Metaanalysen belegt, wie zahlreiche Forschungsüberblicke zeigen (McElvany & Schneider, 2009; Müller, 2015; Philipp, 2012a; Scheerer-Neumann, 2015; Souvignier, 2009). Lesestrategien setzen dort an, wo das Lesen auf Wort- und Satzebene weitestgehend automatisiert ist, wo Kinder aber Probleme haben, das Gelesene auf hierarchiehöherer Prozessebene zusammenzuführen, mit vorhandenem Vorwissen in Beziehung zu setzen und aus Texten zu *lernen* (Gold, Trenk-Hinterberger & Souvignier, 2009; Nix, 2010). Genau auf dieser strategischen und zielgerichteten Ebene des Lesens scheinen leseschwache Kinder oft zu scheitern, und sie unterscheiden sich deutlich von gut lesenden Kindern (Gold et al., 2009; Hartmann, 2006; Philipp, 2012a). Da Lesestrategien zu den effektivsten Fördermassnahmen des Leseverständnisses zählen und im Rahmen von gezielten Förderprogrammen vermittelt werden können, ist es naheliegend, dass Lehrpersonen über ein gewisses Grundlagenwissen zu Lesestrategien verfügen sollten (Philipp, 2015; Scheerer-Neumann, 2015).

Lesestrategien werden in der Regel in vergleichbarer Weise systematisiert, und es wird eine Trennung in zwei Strategiefamilien vorgenommen: 1) die *kognitiven* Verstehensstrategien und 2) die *metakognitiven* Überwachungs- und Kontrollstrategien mit ihren jeweiligen Unterkategorien (Gold, Mokhlesgerami, Rühl, Schreblowski & Souvignier, 2006; Kruse, 2011; Philipp, 2017; Scheerer-Neumann, 2015).

1) Mit *kognitiven Strategien* sind Aktivitäten umschrieben, die sich auf den bewussten, gezielten Umgang mit Texten insbesondere während der Lesephase beziehen (Kruse, 2011). Kognitive Strategien lassen sich in zwei Unterkategorien einteilen: das *Organisieren/Reduzieren* (z.B. Unterstreichen, Überschriften beachten, Zusammenfassungen schreiben, Text graphisch darstellen) und das *Elaborieren/Vorstrukturieren* (z.B. eigenes Wissen aktivieren, Fragen stellen, Wörter klären) (Philipp, 2015; Scheerer-Neumann, 2015).

2) *Metakognitive Strategien* dienen der Überwachung, Steuerung und Regulation der kognitiven Strategien (Philipp, 2012b). Es geht also darum, den Leseprozess *insgesamt* zu überwachen und einen selbstregulierten Umgang mit Lesestrategien zu finden (Gold et al., 2006; Nix, 2010). Wiederum lassen sich hier drei weitere Unterkategorien definieren: die *Planungsstrategien* (Wahl geeigneter Lesestrategien), die *Überwachungsstrategien* (Prozesse zur Überprüfung, ob das Gelesene verstanden wurde) und die *Regulationsstrategien* (Wahl alternativer

Lesestrategien bei Verständnisschwierigkeiten, Lesetempo an Textschwierigkeit anpassen) (Philipp, 2012b; Scheerer-Neumann, 2015).

Eines der einflussreichsten empirisch überprüften Trainingsprogramme zur Förderung des Leseverständnisses ist das „reziproke Lehren“ (Bertschi-Kaufmann & Kappeler, 2010, S. 295; Müller, 2015, S. 26; Nix, 2010, S. 160; Scheerer-Neumann, 2015, S. 107; Seuring, 2010; Souvignier, 2009, S. 194). Es soll hier exemplarisch für weitere effektive Lesestrategietrainings skizziert werden. Das reziproke Lehren besteht aus einer Mischung von kognitiven und metakognitiven Lesestrategien, die meist durch die Lehrperson modellhaft eingeführt werden und darauf abzielen, dass sich Lernende den Text nach und nach selbstverantwortlich erschließen (Bertschi-Kaufmann & Kappeler, 2010). Das Programm geht auf Palincsar und Brown (1984) zurück und beinhaltet vier Strategien: 1) Formulieren von Fragen, 2) Zusammenfassen des Textes, 3) Klären von Wortbedeutungen und Textpassagen, 4) Vorhersagen.

Lesemotivation

Die Lesemotivation ist im Mehrebenenmodell des Lesens von Rosebrock und Nix (2012) auf der Subjektebene angesiedelt und dementsprechend außerhalb hierarchieniedriger oder hierarchiehoher kognitiver Verarbeitungsprozesse einzuordnen. Die Motivation wirkt sich jedoch in übergeordneter Weise auf die Lesekompetenz aus und hat zudem ihren eigenen Wert (McElvany & Schneider, 2009, 161). Mit Bezug auf den Diskurs von Guthrie und Wigfield (2000) umfasst die Lesemotivation nach Philipp (2012c) „individuelle persönliche Ziele, Werte und Überzeugungen hinsichtlich des Themas, Prozesses und Ergebnisses des Lesens“ (S. 31). Die Lesemotivation ist eng verbunden mit dem lesebezogenen Selbstkonzept, d.h. mit der motivationalen Haltung, die sich durch subjektive Erfahrungen und Leistungseinschätzungen längerfristig habituell entwickeln (McElvany, Kortenbruck & Becker, 2008; Nix, 2010). Die Annahme zum Zusammenhang von Lesemotivation und Lesekompetenz bezieht sich darauf, dass sich motivierte Kinder intensiver und ausdauernder mit Texten befassen und umfassendere Lesestrategien anwenden (McElvany & Schneider, 2009). Daneben steht die Lesemotivation in einem Zusammenhang mit einer größeren Lesemenge, was sich wiederum auf das Vorwissen, das positive Selbstkonzept sowie die Leseeffektivität (automatisierte Leseprozesse) auswirkt (McElvany & Schneider, 2009; Philipp, 2012c). Die Lesemotivation selbst lässt sich allerdings durch interessegeleiteten Unterricht (reichhaltige Buch- und Medienangebote) auf der Primarstufe (dritte und vierte Klasse) kaum steigern; zudem hat der interessegeleitete Unterricht auf dieser Schulstufe kaum Wirkung auf die Kompetenzen (Bertschi-Kaufmann & Schneider, 2006). Es wird vermutet, dass der Gewinn des interessegeleiteten Lesens beim frühen

Lesenlernen deshalb gering ist, weil der Fokus in dieser Phase auf dem Erwerb der Basisfertigkeiten liegt, wohingegen ältere Kinder „die Anregung zur individuellen Lektüre für die Lernentwicklung offensichtlich besser verwerten“ (ebd., S. 416). Ergebnisse einer Berliner Längsschnittuntersuchung (dritte bis sechste Klasse) schließen an diese Befunde an und zeigen, dass der Effekt der Motivation auf die Kompetenz (operationalisiert über das Textverständnis) klein ist und dass die Motivation keinen direkten Einfluss auf die mittelfristige Entwicklung von Leseverhalten und Lesekompetenz zwischen vierter und sechster Klasse hat (McElvany et al., 2008). Umgekehrt hat die Lesekompetenz in der dritten Klasse aber einen prädiktiven Einfluss auf die Motivation in der sechsten Klasse. Insgesamt wird die Lesekompetenz von der Motivation und dem Leseverhalten in eher geringem Umfang vorhergesagt, während die beiden Konstrukte (Lesemotivation und Leseverhalten) von der Lesekompetenz beeinflusst werden und zudem untereinander interagieren (ebd.). Auch Ergebnisse der Studie von Gold (2009) zeigen vergleichsweise geringe Einflüsse der Lesemotivation auf die Leseflüssigkeit und das Textverständnis.

Aus der Erkenntnis, dass sich durch die Förderung der Lesemotivation allein keine Verbesserung der Lesekompetenz erreichen lässt – insbesondere nicht bei leseschwachen Kindern – und vermutlich gut ausgebildete basale Lesekompetenzen die Lesemotivation überhaupt erst ermöglichen, ergibt sich eine kritische Betrachtung von Viellese-Verfahren und Leseanimations-Verfahren (Rosebrock & Nix, 2012). Viellese-Verfahren (z.B. stilles Lesen in einem selbstgewählten Buch) zielen darauf ab, insbesondere schwachen und buchfernen Lesenden durch interessante Lektüre eine habituelle Leserrolle nahezubringen (ebd.). Bei Leseanimationsverfahren geht es darum, durch Lesenächte, Lesungen, Bibliotheksbesuche usw. die Verlockung zum Lesen zu fördern und die Motivation zum späteren Lesen zu etablieren (ebd.). Leseanimationsverfahren sind zwar nicht gleichzusetzen mit Viellese-Verfahren, haben aber mit diesen gemeinsam, dass sie nicht direkt auf den Leseprozess (basale Lesekompetenzen) einwirken und zudem eine habituelle, subjektive Bindung zum Lesen bereits *voraussetzen*, damit sie wirksam sind (ebd.). Mit Blick auf die Gruppe leseschwacher und buchferner Kinder verfehlen sie ihre Wirkung deshalb oft, weil im Falle des stillen Lesens die basalen Voraussetzungen fehlen. Diese Kinder lesen gar nicht, wenn sie sollen, sie suchen sich zu schwierige Bücher aus, korrigieren ihre Fehler nicht, und es fehlen ihnen die notwendige Lesestrategien zur Herstellung von Kohärenzen (Herné & Löffler, 2014; Rosebrock & Nix, 2012). Auch Leseanimationsverfahren verfehlen ihre Wirkung oft, bevor die basalen Lesefertigkeiten nicht hinreichend ausgebildet sind (Rosebrock & Nix, 2012; Tacke, 2011).

2.4 Struktur und Merkmale der deutschen Schriftsprache

Um die Leseanforderungen von Leseanfängerinnen und -anfängern zu verstehen, müssen Lehrpersonen zumindest ein Basiswissen zu grundlegenden Merkmalen des deutschen Schriftsystems haben (Schneider, 2017). Dies ist insofern von Bedeutung, weil der Aufbau des deutschen Schriftsprachsystems unterrichtspraktische Konsequenzen für den Schriftspracherwerb hat und weil die Kenntnisse darüber die Lehrpersonen für die Schwierigkeiten, mit denen Kinder mit Leseschwierigkeiten konfrontiert sind, sensibilisieren (Kirschhock, 2004; Topsch, 2005). Im Folgenden wird auf drei Aspekte eingegangen, die für den Anfangsleseunterricht relevant sind und die zudem einen Beitrag zum Verständnis von Leseschwierigkeiten leisten. Zum einen wird die *Graphem-Phonem-Beziehung* beschrieben. Zum anderen wird ein Blick auf die Funktion von *Anlauttabellen* geworfen. Anlauttabellen bauen auf dieser Beziehung auf und gehören zum Repertoire des Anfangsleseunterrichts (Schröder-Lenzen, 2013). Im Weiteren wird die nächsthöhere sprachliche Einheit, die *Silbe*, betrachtet.

Graphem-Phonem-Korrespondenz und Phonem-Graphem-Korrespondenz

Die deutsche Schrift gehört zu den Gruppen der phonographischen Schriften, d.h. lautliche Eigenschaften der gesprochenen Sprache werden verschriftet (Scheerer-Neumann, 2015). Die Korrespondenz zwischen gesprochener und geschriebener Sprache besteht dabei aber nicht auf der Ebene der Laute und Buchstaben, sondern auf der Ebene von Phonemen und Graphemen (ebd.). Phoneme werden als *abstrakte* bedeutungsunterscheidende Einheiten der Lautsprache bezeichnet, wohingegen Grapheme als Buchstaben oder Buchstabenverbindungen (z.B. <sch>, <ch>, <ph> <st>), die mit einem Phonem korrespondieren, beschrieben sind (Marx, 2007b; Schneider, 2017). Diese in den meisten Lehrbüchern gebräuchliche Definition basiert auf der schriftsprachlichen *Funktion* von Phonemen und Graphemen (Mayer, 2016; Scheerer-Neumann, 2018). Entscheidend für diese funktionsbezogene Definition ist also nicht der konkrete, hörbare Sprachlaut (= Phon), sondern die abstrakte Lauteinheit, die – egal wie sie ausgesprochen wird – zu einer Bedeutungsveränderung des Wortes führt (Schröder-Lenzen, 2013, S. 16). Die Ermittlung von Graphemen lässt sich folglich in Analogie zur Ermittlung von Phonemen über Minimalpaare bestimmen wie etwa in den Wörtern <Matte> und <Watte>, die sich nur in den Phonemen /M/ und /W/ unterscheiden, aber zu völlig unterschiedlichen Bedeutungen führen (ebd.).

Eine erste praktische Konsequenz ergibt sich daraus, dass es sich bei der deutschen Schriftsprache nicht um eine rein lauttreue Sprache handelt, bei der Wörter Buchstabe um Buchstabe gelesen werden können, sondern eher um eine „*lautorientierte*“ bzw. „*phonemorientierte*“

Alphabetschrift (Kirschhock, 2004, S. 20; Schneider, 2017, S. 16; Schröder-Lenzen, 2013, S. 16). Unsere gesprochene Sprache enthält nämlich eine weitaus größere Anzahl an bedeutungsunterscheidenden Phonemen (ca. 40) als dies die 26 bzw. 30 Buchstaben, wenn man die Umlaute ä, ö, ü und ß mitzählt, repräsentieren (Metze, 2002; Thomé, 2017). Eine weitere Konsequenz für den Leseprozess ergibt sich aus einer gewissen Asymmetrie und Unregelmäßigkeit in den Graphem-Phonem-Zuordnungsregeln: Im Deutschen gehen wir bei 40 unterschiedlichen Phonemen von 31 verschiedenen Graphemen bzw. Graphem-Clustern wie etwa <sch> aus (Mayer, 2016; Schneider, 2017). So wird das Graphem <e> allein im Wort <Ente> auf lautlicher Ebene unterschiedlich realisiert, einmal als langes /e:/ und am Schluss des Wortes als Schwa-Laut, der nur noch schwach artikuliert wird (Schröder-Lenzen, 2013). Dies bezeichnet die phonetische Mehrdeutigkeit von Graphemen (Kirschhock, 2004). Auf der anderen Seite kann dasselbe Phonem, z.B. das /e/ im Wort <Leben>, <Lehm> und <Meer>, durch ein jeweils anderes Graphem <e> bzw. Orthographem⁷ <eh>, <ee> repräsentiert sein und bezeichnet die graphemische Mehrdeutigkeit (ebd.). Die Regelmäßigkeit ist für die Graphem-Phonem-Korrespondenz (GPK) allerdings größer als für die Phonem-Graphem-Korrespondenz (PGK) (Mayer, 2016). Da beim Lesen die Grapheme in Phoneme übersetzt werden müssen und die Phonem-Alternativen, die für ein Graphem in Frage kommen, überschaubar ausfallen, ist der Leseprozess etwas einfacher als der Schreibprozess (Kirschhock, 2004; Winkes, 2014). Außerdem können geübte Lesende oft auf einfache Wiedererkennungsprozesse von Buchstabenkonfigurationen zurückgreifen, um ein Wort zu erkennen (Kirschhock, 2004). Es kann für die Graphem-Phonem-Zuordnung also von einer höheren „Lauttreue“ gesprochen werden. Umgekehrt gibt es bei der Übersetzung von Phonemen in Grapheme meist eine recht hohe Anzahl Alternativen, wozu orthographisches Wissen notwendig ist (Marx, 2007b, S. 26). Dennoch besteht durch die phonetische Mehrdeutigkeit auch für den Leseprozess eine Hürde, die von Leseanfängerinnen und -anfängern überwunden werden muss. Für eine vertiefte Auseinandersetzung zum Unterschied von Lese- und Schreiberwerb sei hier auf Winkes (2014) verwiesen.

Trotz dieser weiter oben besprochenen Asymmetrie zählt die deutsche Schriftsprache – wie auch die spanische, die italienische oder die finnische – im Unterschied zum Englischen oder Französischen (opakes Schriftsystem) dennoch zu den eher konsistenten bzw. lautgetreuen Orthographien, zumindest was das Lesen betrifft (Mayer, 2013; Schneider, 2017). Durch diese – trotz Einschränkung – hohe lautgetreue Schreibung macht es deshalb durchaus Sinn, Kindern im Anfangsunterricht das phonologische Prinzip zu vermitteln (Mayer, 2016). So zeigt eine

⁷ Orthographem = Statisch seltenere Graphemformen, die bereits orthographische Besonderheiten enthalten (z.B. Doppelungen, Dehnungen) (Schröder-Lenzen, 2013).

empirische Studie von Schabmann, Schmidt, Klicpera, Gasteiger-Klicpera und Klingebiel (2009) folgenden Befund: Je stärker im Unterricht die Betonung auf Buchstaben-Laut-Zuordnungen gelegt wurde, umso eher schienen Kinder in der Lage zu sein, schon sehr früh beim Lesen die vorherrschenden Regeln zu nützen. Dieser buchstabenorientierte Zugang wird aber insbesondere im Zusammenhang mit dem Hilfsmittel „Anlauttabellen“ auch kritisch betrachtet, da durch eine ungünstige Wahl der Bilder eine Phonem-Graphem-Korrespondenz suggeriert wird, die nicht den sprachstatistischen Gegebenheiten entspricht (Schröder-Lenzen, 2013). Im Folgenden wird ein kurzer Blick auf die Bedeutung und die Grenzen von Anlauttabellen geworfen.

Anlauttabellen

Das Potenzial von Anlauttabellen besteht darin, dass sie Kindern im Anfangsunterricht eine Art Gesamtüberblick über die für das deutsche Schriftsprachsystem relevanten Buchstaben-Laut-Beziehungen geben (Riegler, 2010). Allerdings wird die Mehrzahl der auf dem Markt befindlichen Anlauttabellen laut Riegler ihrem komplexen Gegenstand nicht gerecht. Um die Tücken und Grenzen dieses Arbeitsmittels einschätzen zu können, brauchen Lehrpersonen deshalb ein gewisses phonologisches und graphematisches Wissen (ebd.).

Ein erstes Problem für den Einsatz von Anlauttabellen beim Leseerwerb ergibt sich aus der besprochenen Varianz der Graphem-Phonem-Korrespondenz, da ein Graphem verschiedene Phoneme repräsentieren kann wie z.B. das Graphem <V> in den Wörtern <Vogel> und <Vase>, das lautlich einmal als /f/ und einmal als /w/ realisiert wird (Schröder-Lenzen, 2013, S. 25). Dennoch wäre ein Graphem-basierter Erstleseunterricht geeigneter als ein buchstabenorientierter Unterricht (Thomé, 2017). Begründet wird dies mit dem eben besprochenen günstigeren Verhältnis von Graphem-Phonem-Korrespondenz im Gegensatz zum Buchstaben-Laut-Verhältnis (ebd.). Auch die Analyse von Phonemen bleibt dabei ein primär kognitiver Akt und kein auditiver, da auch die schriftlichen Realisierungen von Lauten in einem Wort oft nicht ohne weiteres zu *hören* sind (Schröder-Lenzen, 2013).

Eine weitere Kritik zum Einsatz von Anlauttabellen hängt mit der Häufigkeit des Auftretens sogenannter Basis- und Orthographeme zusammen, die gemäß Thomé in zahlreichen Anlauttabellen irreführend eingeführt werden. Ein Basisgraphem bezeichnet die häufigste Schreibung bzw. die dominante Verschriftung, die mit einem bestimmten Laut korrespondiert. So ist für den langen i-Laut /i:/ <ie> die Basis wie z.B. im Wort <die>. Entsprechend bezeichnet das Orthographem alle seltener auftretenden Schreibungen wie z.B. das <ih> im Wort <ihre>, das <i> im Wort <dir> oder das <ieh> im Wort <ziehen> (Thomé, 2017). Vor diesem Hintergrund werden zwei häufige Probleme von Anlauttabellen offensichtlich. Zum einen zeigt sich, dass

Anlautbilder und Begriffe, die übliche Verschriftungsformen repräsentieren sollten, nicht in ihrer *dominanten* Verschriftungsform veranschaulicht und dass zum anderen oft falsche Zuordnungen von Bild und Buchstabenlautung gewählt werden (Schründer-Lenzen, 2013). Dies lässt sich mit dem vielbeschriebenen „Igel-Fehler“ zeigen (Corvacho del Toro, 2013, S. 142; Füssenich & Löffler, 2008, S. 76). So eignet sich das Wort <Igel> entgegen der häufigen Verwendung in Anlauttabellen und anderen Lernmaterialien nicht, um das Graphem <i> als Korrespondenz zum langen Phonem /i:/ einzuführen. Hier ist der Anlaut (langes /i:/ im Wort <Igel>) und das Graphem <I> schlicht falsch zugeordnet. Was bei diesem Beispiel außerdem sichtbar wird: In Anlauttabellen werden häufig relativ selten vorkommende Orthographeme eingeführt (z.B. das <i>) und nicht die weitaus häufigeren Basisgrapheme, z.B. das <ie> (Thomé, 2017).

Neben diesen linguistischen Kriterien zur Beurteilung von Anlauttabellen sind weitere phonetische, semantische und graphische Kriterien zu berücksichtigen, die Mayer (2016) auf der Grundlage von Forschungsarbeiten von Kruth und Thul (2003) sowie Crämer, Füssenich und Schumann (1996) vorschlägt:

- 1) Auf phonetischer Ebene sind kürzere ein- oder zweisilbige Wörter, die zudem keine Konsonantenhäufungen (insbesondere am Wortanfang) enthalten, geeigneter als lange Wörter mit komplexer Vokal-Konsonanten-Struktur.
- 2) Wörter, deren erste Silbe dem Buchstabennamen entspricht (<Kamel>, <Besen>) sollten vermieden werden.
- 3) Mit Bezug auf semantische Kriterien eignen sich Wörter bzw. Bilder, die der Erfahrungswelt von Kindern entsprechen und zudem nur eine Wortassoziation zulassen.
- 4) Auf graphischer Ebene schließlich ist auf Bildmaterial zu achten, das optisch ansprechend ist und eindeutig mit dem gemeinten Wort assoziiert wird.

Silben und Silbenbögen

Aus didaktischer Perspektive kann es sinnvoll sein, den Fokus nicht nur auf die kleinste lautliche Einheit (Phonem) von Wörtern zu legen, sondern auf die nächstgrößere Einheit, die Silbe. Eine Orientierung am silbischen Prinzip ist aus sprachwissenschaftlicher Perspektive bedeutsam, da dieses u.a. die Prozesse der Worterkennung beim Lesen erleichtert (Ritter, 2005; Röber, 2013; Schründer-Lenzen, 2013). Mit Blick auf den phonologischen Aspekt des Lesens lassen sich Hürden von Leseanfängerinnen und -anfängern damit erklären, dass beim Aussprechen von (gelesenen) Wörtern „die Einheit der Artikulation nicht einzelne Laute, sondern Silben sind“ (Kirschhock, 2004, S. 22). Geübte Lesende nutzen die Silbenstruktur als visuelle und

artikulatorische Einheit aus, was zu einer Entlastung des Arbeitsgedächtnisses führt (Ritter, 2005). Dagegen gehen ungeübte Kinder bzw. Kinder mit Leseschwierigkeiten oft den frustrierenden Umweg über eine Kunstsprache, die beim Lautieren und Synthetisieren von einzelnen Buchstaben entsteht (Röber & Olfert, 2010).

Hier zunächst ein Wort zum regulären Aufbau der Silbe: Jede Silbe verfügt über einen vokalischen Silbenkern (Nukleus) und einen fakultativen konsonantischen Anfangs- (Onset) und Endrand (Coda) (Scheerer-Neumann, 2018). Das typische Betonungsmuster bei zweisilbigen Wörtern ist der Trochäus mit Hauptbetonung auf der ersten Silbe (Stammsilbe) und einer zweiten unbetonten Reduktionssilbe wie z.B. im Wort <Mus-ter> [1. Silbe: M = Onset, U = Nukleus, S = Koda; 2. Silbe: T = Onset, E = Nukleus, R = Koda] (Schröder-Lenzen, 2013, S. 33). Je nach Stellung von Graphemen innerhalb der Silbe (z.B. das <ch> in <mach-en> oder <si-cher>), werden sie anders ausgesprochen. Außerdem spielt die richtige Betonung der Silbe für die Sinnentnahme eine wichtige Rolle (Kirschhock, 2004).

Der Umstand, dass Kinder bereits früh auf die sprachliche Einheit der Silbe aufmerksam werden (z.B. beim Reimen), kann als Argument dienen, Silben didaktisch von Anfang an zu nutzen (Füssenich & Löffler, 2008; Helbig, Kirschhock, Martschinke & Kummer, 2005), denn „Silben sind die kleinste spontan zu äußernde Einheit der Sprache – nicht Laute, nicht Wörter“ (Röber & Olfert, 2010, S. 1). Eine Anwendung findet sich etwa durch die Markierung von Silben durch Silbenbögen oder farbliche Markierungen. Die Grundidee ist dabei, dass durch die visuelle Unterstützung das Erkennen der Wortgliederungsebene (Silbe) erleichtert wird und das „Zusammenschleifen“ von Einzellaute leichter gelingt; außerdem bietet die Silbe eine Annäherung an den Wortakzent des Zielwortes (Schröder-Lenzen, 2013, S. 36). Empirische Belege für die erfolgreiche Anwendung von visuellen Silbensegmentierungsansätzen finden sich in verschiedenen Studien (Ritter, 2005; Ritter & Scheerer-Neumann, 2009; Tacke, 2005).

Nicht nur das Wissen um das Verhältnis von Graphemen und Phonemen, sondern auch das Wissen um die silbische Struktur der Sprache ist entscheidend für den Leseerwerb. Anlautta-bellen und Silbenbögen sind häufige didaktische Mittel, die auf diesem Wissen aufbauen und helfen, die durch linguistische Gegebenheiten entstehenden Hürden zu überwinden. Die voran-gehenden Ausführungen sind zudem zentral für das Verständnis der nachfolgend ausgeführten Vorläuferfertigkeiten des Lesens (Kap. 2.5) und der Leseentwicklung ab der ersten Klasse (Kap. 2.6).

2.5 Vorläuferfertigkeiten des Lesens: Phonologische Informationsverarbeitung

Bereits im Vorschulalter – etwa mit drei bis vier Jahren – erwerben Vorschulkinder erste Einsichten, die für den Schriftspracherwerb bedeutsam sind, wenn auch zunächst oft noch in sehr oberflächliche Weise (Schneider, 2017). Neben der Buchstabenkenntnis, dem Wortschatz und der Vorerfahrung mit Schriftlichkeit („Literacy“) ist der Aspekt der phonologischen Informationsverarbeitung zentral. Mit Blick auf die einschlägige Literatur „wird in diesem Zusammenhang von *spezifischen Vorhersagemerkmale* der Lesekompetenz gesprochen“ (Schneider, 2017, S. 36). Vorläuferfertigkeiten werden in Anlehnung an Torgesen, Wagner, Rashotte, Burgess und Hecht (1997) meist unter dem Konstrukt der „phonologischen Informationsverarbeitung“ zusammengefasst, die sich in drei Teilkomponenten untergliedern lässt: *phonologisches Arbeitsgedächtnis*, *phonologische Bewusstheit* sowie *Zugriffsgeschwindigkeit bzw. Benennungsgeschwindigkeit* (Mayer, 2013). Diese drei Bereiche können sowohl theoretisch begründet wie auch aufgrund einer Vielzahl empirischer Befunde als bedeutendste individuelle Voraussetzungen für den Schriftspracherwerb betrachtet werden (Marx, 2007b).

2.5.1 Phonologisches Arbeitsgedächtnis

Das phonologische Arbeitsgedächtnis spielt im Zusammenhang mit dem Schriftspracherwerb eine zentrale Rolle. In Forschungsübersichten (Götze, Hasselhorn & Kiese-Himmel, 2000; Mayer, 2013) wird die Bedeutung bestimmter phonologischer Gedächtnisleistungen hinsichtlich der späteren Entwicklung der Worterkennung und des Leseverständnisses belegt.

Das Arbeitsgedächtnis setzt sich aus verschiedenen Komponenten zusammen (Modell des Arbeitsgedächtnisses nach Baddeley, 2000, S. 421). Das phonologische Arbeitsgedächtnis ist ein Teilbereich davon und setzt sich seinerseits aus den Elementen *phonetischer Speicher* (phonological buffer) und *subvokaler artikulatorischer Kontrollprozess* (phonological rehearsal) zusammen (Götze et al., 2000, S. 16; Mayer, 2013, S. 38). Ganz allgemein dient das phonologische Arbeitsgedächtnis der Zwischenspeicherung und Weiterverarbeitung sprachlicher Informationen (Mayer, 2013). Die Aufgabe des phonologischen Buffers ist es, auditorisch-verbale Informationen für einen Zeitraum von bis zu zwei Sekunden für die weitere Verarbeitung zur Verfügung zu stellen. Durch das phonologische Rehearsal, d.h. durch eine Art „inneres Sprechen“ (Götze et al., 2000, S. 16) verbleiben Informationen länger im phonetischen Speicher (Mayer, 2016). Daraus lässt sich Folgendes ableiten: Je schneller Wörter bzw. Pseudowörter artikuliert werden, desto mehr Einheiten können innerhalb des Zeitfensters (phonologischer Buffer) wiederholt werden und desto länger ist die Gedächtnisspanne (Mayer, 2013). Hier wird der Bezug zwischen phonologischem Arbeitsgedächtnis und den später dargestellten Prozess- und Stufenmodellen deutlich. Die Herausforderung besteht darin, dass beim alphabetischen

Lesen die Ergebnisse der Graphem-Phonem-Konversionen bis zur Synthese des ganzen Wortes gespeichert werden müssen und dass beim Satz- und Textverstehen schon Gelesenes bis zum endgültigen Verständnis der Aussage verfügbar sein muss (Scheerer-Neumann, 2018). Die erfasste Kapazität des phonologischen Arbeitsgedächtnisses (z.B. Nachsprechen von Pseudowörtern) gilt mit Verweis auf entsprechende empirische Studien als zuverlässiger Prädiktor für die spätere Schriftsprachentwicklung (Worterkennung) (Marx, 2007b; Mayer, 2013; Petermann & Daseking, 2015).

2.5.2 Phonologische Bewusstheit

Der phonologischen Bewusstheit als Teilkomponente der phonologischen Informationsverarbeitung wurde in der wissenschaftlichen Forschung in den letzten Jahren im Vergleich zu den anderen Vorhersagemerkmalen die größte Beachtung geschenkt (Schründer-Lenzen, 2013). Ein Grundverständnis dieser zentralen Vorläuferfertigkeit ist deshalb für Lehrpersonen in verschiedener Hinsicht bedeutsam.

Dimensionen und Entwicklung der phonologischen Bewusstheit

Unter phonologischer Bewusstheit wird allgemein die metakognitive Fähigkeit verstanden, die es ermöglicht, die Aufmerksamkeit auf die lautliche Struktur von Wörtern zu richten und vom Wortsinn zu abstrahieren (Scheerer-Neumann, 2018; Schnitzler, 2008, 5; Schründer-Lenzen, 2013). Im Einzelnen beinhaltet das Konstrukt der phonologischen Bewusstheit die bewusste Identifizierung, Analyse, Synthese und Manipulation sprachlicher Einheiten auf sublexikalischer Ebene (Mayer, 2013; Tacke, 2011). Die wohl bekannteste Systematisierung des Konstrukts liefert Schnitzler (2008) in ihrem Modell zur phonologischen Bewusstheit und ihren umfassenden theoretischen Ausführungen und empirischen Belegen zu seiner Gültigkeit. Zentrales Moment des Modells ist die Berücksichtigung der beiden Dimensionen „phonologische Einheit“ (Silbe, Onset-Rime) und „Explizitheit der Operationen“ (Identifizieren, Synthetisieren, Segmentieren, Manipulieren) (ebd., S. 21f.). Auf der Grundlage des Schnitzler-Modells zeigt die folgende Abbildung 3 von Mayer (2013), wie diese beiden Dimensionen mit entsprechenden Aufgaben inhaltlich gefüllt sind.

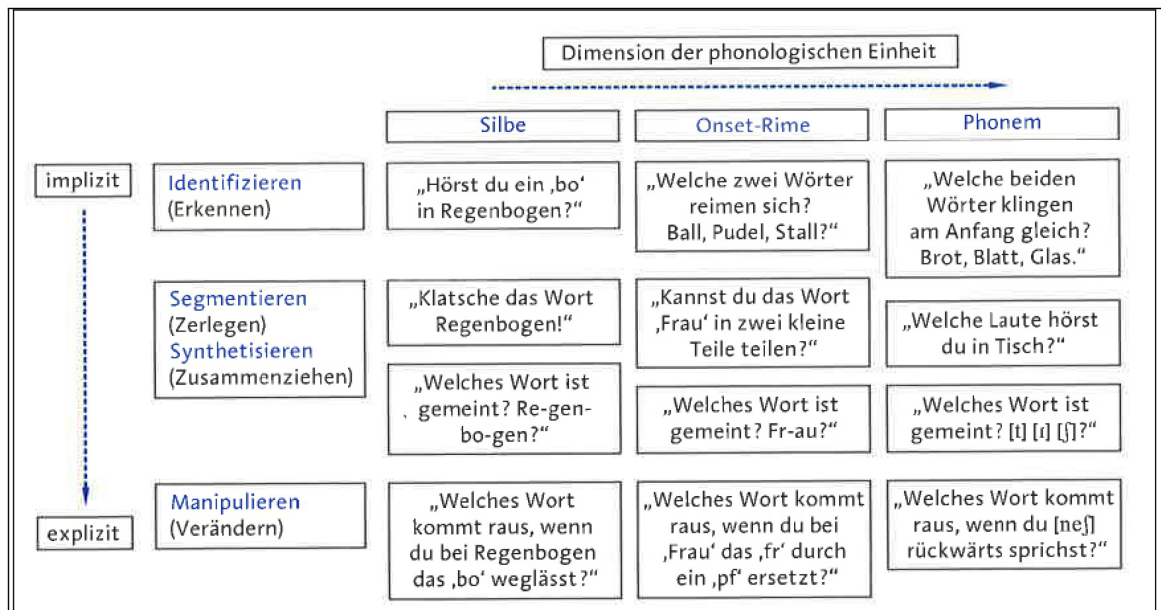


Abbildung 3. Das Zweidimensionale Modell der phonologischen Bewusstheit von Schnitzler (2008) mit Aufgabenbeispielen von Mayer (2013, S. 49).

Unter Einbezug der beiden Dimensionen der phonologischen Bewusstheit lässt sich die Schwierigkeit von Aufgaben ableiten. Sie ist umso komplexer, je kleiner die sublexikalische Einheit (Phonem) und je expliziter die kognitive Operation (Manipulation) ist (Mayer, 2013; Scheerer-Neumann, 2018; Schnitzler, 2008). Unter der Bezeichnung „phonologische Bewusstheit im weiteren Sinne“ werden Aufgaben aufgeführt, die sich in der Regel bereits im Kindergarten entwickeln, sich auf der Ebene von Silben und Onset-Rime befinden und mit impliziten Operationen durchgeführt werden können (Schnitzler, 2008, S. 20). Die „phonologische Bewusstheit im engeren Sinne“ bezeichnet dementsprechend die Fähigkeit, auf der Ebene von Phonemen explizite Operationen durchzuführen (ebd.). Diese kontinuierliche Entwicklung der Schwierigkeiten (Zeitraum Kindergarten bis viertes Schuljahr) von den großen zu den kleinen phonologischen Einheiten und von der impliziten zur expliziten Bewusstheit konnte aufgrund der von Schnitzler (2008) einbezogenen Studien bestätigt werden, auch wenn sie einräumt, dass hierzu weiterer Forschungsbedarf besteht. Grundsätzlich kann aber davon ausgegangen werden, dass Kinder noch vor Schulbeginn Aufgaben im Sinne der phonologischen Bewusstheit im weiteren Sinne lösen können, während sich die Bewusstheit im engeren Sinne in der Regel erst im Zusammenhang mit dem Leselehrgang der ersten Klasse entwickelt (Skowronek & Marx, 1989). Hier stellt sich die Frage, ob die Ausbildung der phonologischen Bewusstheit im engeren Sinne nicht auch die Folge der Erfahrung mit Schrift sein könnte (Marx, 2007b, S. 56; Schneider, 2017, S. 47) und in diesem Sinne keine „Vorläuferfertigkeit“ darstellt. Diese Frage ist aufgrund der nicht eindeutigen Ergebnisse zahlreicher Studien nach wie vor nicht vollständig geklärt (Landerl et al., 2019). Längsschnittdaten deuten „auf eine reziproke Beziehung hin in

der Art, dass die phonemanalytischen Kompetenzen am Schulanfang zwar nicht irrelevant sind, ihre weitere Entwicklung vor allem auf der Phonemebene jedoch in Interaktion mit dem Lesen- und Schreibenlernen erfolgt“ (Scheerer-Neumann, 2018, S. 138). Auch aus den Ergebnissen der Studie von Klicpera und Gasteiger-Klicpera (1993) lässt sich dieser reziproke Einfluss ableiten, da hier aufgezeigt wird, dass in den ersten Schulwochen ein Zugewinn an phonologischer Bewusstheit erfolgt (Marx, 2007b). Damit kann mit Sicherheit gesagt werden, dass die Konfrontation mit Schriftsprache im Anfangsunterricht einen wesentlichen Anteil an der Herausbildung der phonologischen Bewusstheit hat (Klicpera et al., 2017).

Bedeutung der phonologischen Bewusstheit für spätere Lesekompetenzen

Eine kritische Betrachtung zum aktuellen Forschungsstand zeigt, dass die Ergebnisse verschiedener Studien zur phonologischen Bewusstheit in Bezug auf ihren prädiktiven Wert für die Lesekompetenz ein unklares Bild zeichnen (siehe Forschungsüberblicke: Klicpera et al., 2017; Marx, 2007b; Scheerer-Neumann, 2018; Schneider, 2017; Valtin, 2010). Unklar bleibt vor allem die bereits angesprochene kausale Rolle der phonologischen Bewusstheit (Schneider, 2017; Valtin, 2010). Für die Interpretation des Vorhersagewertes der phonologischen Bewusstheit sind zum einen Langzeitstudien interessant, die den Zusammenhang zwischen Prädiktorvariablen (phonologische Bewusstheit) und Kriteriumsvariablen (Leistungen im Lesen am Ende der ersten Klasse oder später / Graphem-Phonem-Korrespondenz / alphabetisches Rekodieren) untersuchen (Ennemoser, Marx & Schneider, 2012; Landerl & Wimmer, 2008; Wimmer, Landerl, Linortner & Hummer, 1991). Zum anderen interessieren auch Interventionsstudien, bei denen die phonologische Bewusstheit aktiv gefördert wird und die spätere Leseleistungen darauf bezogen bewertet werden.

Interventionsstudien bzw. Präventionsstudien (u.a. Hartmann, 2002; Küspert & Schneider, 2008; Souvignier, Duzy & Schneider, 2012) weisen darauf hin, dass sich Effekte des Trainings zur phonologischen Bewusstheit vor allem bei der phonologischen Bewusstheit selbst sowie bei frühen Lesefertigkeiten (Dekodieren) nachweisen lassen. Die Wirkung nimmt mit zunehmendem Alter ab, und der Zusammenhang mit hierarchiehöheren Prozessen wie dem Leseverständnis zeigt sich kaum mehr (Ehri et al., 2001; Huemer, Pointner, Schöfl & Landerl, 2019; Ise, Engel & Schulte-Körne, 2012; Mayer, 2016; Schneider et al., 2013; Valtin, 2010). Der abnehmende Zusammenhang zwischen phonologischer Bewusstheit und der späteren Leseentwicklung lässt sich durch das reziproke Verhältnis zwischen phonologischer Bewusstheit und Schriftspracherwerb erklären. Galuschka und Schulte-Körne (2015) gehen davon aus „dass der Schriftspracherwerb ein umfassendes Verständnis für die Phonemstruktur der Sprache

ermöglicht und damit eine Förderung der phonologischen Bewusstheit zu keiner weiteren, bedeutsamen Verbesserung der Leseleistung führt“ (S. 482). Auch Korrelationsstudien aus dem deutschsprachigen Raum (u.a. Berendes, Schnitzler, Willmes & Huber, 2010; Ennemoser et al., 2012; Gorecki & Landerl, 2015; Landerl & Wimmer, 2008; Wimmer et al., 1991), die den Vorhersagewert der phonologischen Bewusstheit u.a. hinsichtlich basaler Lesekompetenzen erfassen, verweisen darauf, dass die vorschulische phonologische Bewusstheit vor allem in engerer Beziehung zur Rechtschreibleistung (Ende erstes und viertes Schuljahr), aber weniger in Beziehung zu späteren Lesekompetenzen (Lesegeschwindigkeit) steht (Klicpera et al., 2017; Mayer, 2016; Scheerer-Neumann, 2018). Weiter ist davon auszugehen, dass die Leseleistungen, die durch die phonologische Bewusstheit vorausgesagt werden, vor allem die phonologischen Dekodierleistungen etwa am Ende des ersten Schuljahres (alphabetische Phase) betreffen, weshalb sie demnach indirekt wirkt (siehe Studien Ennemoser et al., 2012; Wimmer et al., 1991). Somit scheinen andere Prädiktoren wie die Benennungsgeschwindigkeit oder der Umfang des Wortschatzes in einem bedeutsameren Zusammenhang mit der Leseleistung (Lesegeschwindigkeit) zu stehen (Mayer, 2016; Scheerer-Neumann, 2018).

Gemäß Forschungsüberblick von Scheerer-Neumann (2018) zeigt sich, dass die *positive* Leseentwicklung bei Kindern mit guten phonologischen Fähigkeiten zuverlässiger vorausgesagt werden kann als bei vorschulisch schwächeren Kindern. Offensichtlich können jedoch Kinder mit ungünstigen Voraussetzungen ihre Schwierigkeiten im Erstleseunterricht teilweise kompensieren. Damit sind vorschulische Defizite in der phonologischen Bewusstheit kein zwingender Risikofaktor für Lese-Rechtschreibschwierigkeiten (Mayer, 2016). Bei einem Teil der Kinder scheint eine geringe vorschulische phonologische Bewusstheit allerdings durchaus ein Risikofaktor zu sein (Scheerer-Neumann, 2018).

2.5.3 Benennungsgeschwindigkeit

Als dritte Komponente der phonologischen Informationsverarbeitung gilt die Fähigkeit, eine Abfolge von vertrauten Bildern und Symbolen (z.B. Zahlen) möglichst schnell zu identifizieren, die entsprechenden verbalen Repräsentationen im mentalen Lexikon zu aktivieren und das Wort schließlich zu artikulieren (Mayer, 2016, 99). Synonym für die Benennungsgeschwindigkeit wird auch der Begriff „Rapid Automatized Naming“ (RAN) oder „Naming-Speed“ verwendet (Scheerer-Neumann, 2018, S. 139). Auf Grundlage empirischer Studien lässt sich für die Benennungsgeschwindigkeit übereinstimmend sagen, dass es sich um einen relativ stabilen Prädiktor für das Lesen (Lesegeschwindigkeit, Satz- und Textverständnis) handelt (Ennemoser et al., 2012; Landerl & Wimmer, 2008). Wie aus der Ländervergleichsstudie von Landerl et al. (2019) hervorgeht, scheint RAN im Gegensatz zur phonologischen Bewusstheit ein von der jeweiligen

Orthographie unabhängiger Prädiktor für die Leseflüssigkeit zu sein. Die Autorinnen und Autoren sprechen deshalb sogar von einem „language-universal cognitive mechanism“ (ebd., S. 220).

Überprüft wird RAN üblicherweise, indem den Versuchspersonen altersabhängig unterschiedliche Symbole (Farben, bekannte Objekte, Buchstaben, Zahlen) vorgelegt werden, die in Lese- richtung so rasch wie möglich benannt werden müssen (Mayer, 2016). Allerdings herrscht im wissenschaftlichen Diskurs Uneinigkeit darüber, was damit *genau* gemessen wird und wie der Zusammenhang zwischen RAN und dem Lesen theoretisch zu erklären ist (Klicpera et al., 2017; Mayer, 2016). Dabei können mit Bezug auf die Übersichtsarbeiten von Kirby et al. (2010) und Papadopoulos, Spanoudis und Georgiou (2016) mindestens drei theoretische Stränge verfolgt werden, die sich aber nicht ausschließen müssen, sondern jeweils zur Erklärung des Konstrukts RAN und dessen Zusammenhang auf die Lesekompetenz beitragen.

- 1) Zum einen wird die theoretische Nähe des Konstrukts RAN zur phonologischen Bewusstheit diskutiert (Kirby et al., 2010; Mayer, 2016). Übereinstimmung besteht darin, dass RAN zwar eine „phonologische Komponente“ besitzt (Mayer, 2016, S. 106), dass die phonologische Bewusstheit jedoch nur moderat mit der Benennungsgeschwindigkeit zusammenhängt. Die beiden Konstrukte haben auch einen jeweils unabhängigen Erklärungswert für die Lesekompetenz (Kirby et al., 2010; Klicpera et al., 2017; Papadopoulos et al., 2016; Wimmer, 1993).
- 2) Ein zweiter Erklärungsansatz zum Zusammenhang von RAN und Schriftspracherwerb betrifft die globale kognitive Verarbeitungsgeschwindigkeit, die sich mit zunehmendem Alter mehr und mehr entwickelt (Mayer, 2016). In diesem Sinne würde das Konstrukt eher in Zusammenhang mit Prozessen des Arbeitsgedächtnisses stehen, d.h. dass Kinder orthographische und phonologische Repräsentationen weniger schnell aufeinander beziehen können, was sich in der Folge auch auf die Lesegeschwindigkeit auswirkt (Papadopoulos et al., 2016).
- 3) Eine dritte Vermutung bezieht sich darauf, dass die Benennungsgeschwindigkeit ein Maß dafür darstellt, wie effektiv orthographische Repräsentationen im Langzeitgedächtnis abstrahiert werden können (Mayer, 2016). Mit der Benennungsgeschwindigkeit würde mit Bezug zum Zwei-Wege-Modell des Lesens (Kap. 2.3.2) die orthographische Verarbeitungseffizienz überprüft, die sich beim Lesen von echten Wörtern eher zeigen müsste als beim Lesen von Pseudowörtern. Die Hypothese beruht auf dem in zahlreichen Studien festgestellten Zusammenhang zwischen Benennungsgeschwindigkeit und Leseflüssigkeit bzw. automatisierter Worterkennung, die mit dem Erfassen größerer sprachlicher Einheiten einhergeht. Während die phonologische Bewusstheit eher mit dem phonologischen Rekodieren bzw. indirekten Worterkennen in Verbindung steht, ist die Benennungsgeschwindigkeit eher mit Prozessen des automatisierten Lesens

jenseits des ersten Schuljahrs und damit mit späteren Phasen des Leseerwerbs assoziiert (Klicpera et al., 2017; Mayer, 2016; Scheerer-Neumann, 2018). Im deutschsprachigen Raum finden sich empirische Belege dazu u.a. bei Ennemoser et al. (2012) sowie bei Landerl und Wimmer (2008).

Gemäß einer Übersichts-Studie von Papadopoulos et al. (2016) ist davon auszugehen, dass die phonologische Bewusstheit und RAN je eigene Erklärungen der Varianz des lauten Lesens (nicht aber des stillen Lesens) leisten. Im Rahmen der Früherkennung von Kindern mit Risiken im Leseerwerb scheint es deshalb sinnvoll, neben der phonologischen Bewusstheit ergänzend auch die Benennungsgeschwindigkeit zu prüfen, um zu zuverlässigeren Vorhersagen zu kommen.

2.6 Leseentwicklung: Entwicklungs- und Stufenmodelle

Im folgenden Kapitel werden die Grundlagen der Leseentwicklung, wie sie in *Stufenmodellen* beschrieben sind – und wie sie sich etwa bis Ende der Primaschulzeit gestalten – vorgestellt.

In Entwicklungsmodellen werden Parallelitäten der Lese- und Rechtschreibentwicklung immer wieder betont, und ihr gegenseitiger Einfluss wird verdeutlicht (Klicpera et al., 2013). Die Prozesse des Lesens und Schreibens hängen zweifelsohne zusammen, können aber trotz aller Gemeinsamkeiten nicht als rein symmetrische Prozesse verstanden werden (Winkes, 2014). Der offensichtlichste Unterschied lässt sich in den Worten von Winkes (2014) so zusammenfassen, „dass korrektes Schreiben schwieriger ist als fehlerfreies Lesen, da beim Schreiben die komplette Grapheminformation zur Verfügung stehen muss. Beim Lesen hingegen handelt es sich um einen Wiedererkennungsprozess, der auch mit weniger gutem Wissen über das Wort gelingen kann“ (S. 150). In den folgenden Ausführungen wird ausschließlich auf die lesebezogene Seite fokussiert, im Wissen darum, dass die Modelle oft beide Seiten berücksichtigen.

Stufenmodelle auf Grundlage von Informationsverarbeitungstheorien

In der psychologischen Leseforschung haben sich insbesondere Stufenmodelle durchgesetzt, welche die Entwicklung visueller Worterkennung als Abfolge unterscheidbarer Phasen beschreiben (Richter & Müller, 2017). Modelle, die sich auf kognitive Informationsverarbeitungstheorien beziehen, beschreiben die basale Leseentwicklung auf Wortebene und lassen sich mit Modellen wie etwa dem Zwei-Wege-Modell (Kap. 2.3.2) vereinbaren (Klicpera et al., 2013). Die theoretischen Annahmen basieren auf Konzepten, wonach nicht-lexikalisches Lesen (indirekter Zugriff auf Wörter mittels phonologischem Kodieren) sukzessive vom lexikalischen Lesen (direkter Zugriff auf Wörter mittels mentalem Lexikon) abgelöst wird (Helbig et al., 2005;

Klicpera et al., 2013). Die bekanntesten Stufenmodelle (auch als Phasenmodelle oder Strategiemodelle bezeichnet) gehen in ihren Kernannahmen auf Frith (1985) und Ehri (1997) zurück. Sie wurden im Verlaufe der Zeit weiterentwickelt und auf den deutschen Sprachraum adaptiert, z.B. von Günther (1986) sowie Scheerer-Neumann (2006). Allen gemeinsam ist die Grundannahme einer stufenweisen Entwicklung des Lese- bzw. Schreiberwerbs von der *präliteral-symbolischen*, *logographischen* über die *alphabetische* zur *orthographischen* Phase (Klicpera et al., 2013, S. 30), auch wenn sich die Modelle in der Terminologie und Differenziertheit teilweise geringfügig unterscheiden (Mayer, 2013). Empirische Studien belegen die Grundannahmen von Stufenmodellen sowohl für den angloamerikanischen Raum als auch mit Einschränkungen für den deutschsprachigen Raum (Kirschhock, 2004). Die Einschränkung bezieht sich vor allem auf die logographemische Phase, die im deutschen Sprachraum von wesentlich geringerer Bedeutung ist als im englischsprachigen Raum (ebd.). Die zweite Kritik bezieht sich auf die Abgrenzbarkeit einzelner Phasen. Die Modelle wurden deshalb dahingehend weiterentwickelt, dass nach aktuellem Kenntnisstand nicht mehr von in sich abgeschlossenen Phasen auszugehen ist, sondern von „dominanten Strategien“ (Helbig et al., 2005, S. 41; Kirschhock, 2004, S. 47) und einer zunehmenden Integration verschiedener Teilfertigkeiten, die sich parallel oder überlappend entwickeln können (Mayer, 2016). Letzteres wird etwa im weiterentwickelten Kompetenzentwicklungsmodell (Abbildung 4) von Klicpera et al. (2017) oder im differenzierten Entwicklungsmodell der dominanten Lesestrategien von Helbig et al. (2005) deutlich.

Abbildung 4. Kompetenzentwicklungsmodell für das Worterkennen und laute Lesen (Klicpera et al., 2017, S. 29)

hilfreich (Klicpera et al., 2017; Marx, 2007b). Nur auf dieser Basis können bestimmte „Fehler“ als notwendige Entwicklungsschritte interpretiert werden (Füssenich & Löffler, 2008, S. 73; Kirschhock, 2004, S. 44). In praktischer und diagnostischer Hinsicht „scheint eine genaue Analyse der vom Kind verwendeten Strategien und Zugänge zum Lesen und Rechtschreiben von Bedeutung“ (Klicpera et al., 2017, S. 42), weshalb die Stufen bzw. Strategien im Folgenden als getrennte Phasen beschrieben werden, im Wissen darum, dass die Übergänge fließend sind.

Präliteral-symbolische und logographemische Strategie: Nicht-lexikalisches Lesen

Die präliteral-symbolische Phase setzt meist bereits vor dem Schulalter ein und bezeichnet ganz allgemein die Phase, in der Kinder die Erfahrung machen, dass die dreidimensionale Realität durch Abstraktion bzw. Symbolik in eine zweidimensionale Darstellung überführt werden kann (Füssenich & Löffler, 2008; Mayer, 2016). Dies zeigt sich etwa darin, dass Kinder „so tun, als ob“ sie lesen würden und Bilderbücher „vorlesen“, indem sie eine deutlich ausgeprägte Prosodie wählen und auf grammatikalisch vollständige Sätze achten (Mayer, 2016, S. 39).

In der logographemischen Phase lernen die Kinder, ein Wort aufgrund einzelner visueller Merkmale oder bestimmter Kontexte (Logos, Embleme, Produktnamen) oder aufgrund einzelner Buchstaben (z.B. ihren Namen) zu erkennen (Richter & Müller, 2017). Der Fortschritt besteht darin, dass Kinder erfassen, „dass die Symbole der Schrift Sprache abbilden“ (Mayer, 2016, S. 40). Die Zuordnung von Graphemen zu Phonemen findet in dieser Phase aber noch nicht systematisch statt, so dass auch die Reihenfolge der Grapheme sekundär ist (Füssenich & Löffler, 2008; Scheerer-Neumann, 2015). Es bestehen allerdings Zweifel darüber, ob diese Art der Worterkennung von allen Kindern zwingend durchlaufen wird. Dies gilt insbesondere in Sprachräumen mit einigermaßen lautgetreuen Schriftsprachsystemen wie dies im Deutschen etwa der Fall ist (Helbig et al., 2005; Klicpera et al., 2017; Mayer, 2016; Richter & Müller, 2017). Die Studie von Wimmer und Hummer (1990) zeigt, dass bei deutschsprechenden Erstklässlern keine Zeichen der Anwendung der logographemischen Strategie zu finden waren. Das Auftreten dieser Phase scheint gemäß Klicpera et al. (2013) zwar möglich, aber nicht zwingend zu sein für die nachfolgende Phase der Leseentwicklung. Sie zeigt sich zudem meist nur in einem kurzen Zeitraum und hängt in erster Linie von der Art der Erstleseinstruktion ab. Im deutschsprachigen Raum wird der Fokus im Erstleseunterricht in der Regel auf die Einführung der Graphem-Phonem-Zuordnung gelegt, was damit begründet wird, dass es sich beim Deutschen um eine im Vergleich zum Englischen relativ konsistente phonologische Schriftsprache handelt (Marx, 2007b).

Alphabetische Strategie: Phonologisches Rekodieren auf sublexikalischer Ebene

In dieser Phase beginnt das Lesen und Schreiben erst richtig, indem die Buchstaben-Lautbeziehung unserer lautorientierten Sprache umgesetzt wird (Günther, 2007). Das Erlesen von Wörtern erfolgt nun über das regelhafte phonologische Rekodieren der Buchstabenreihe, wie dies im Zwei-Wege-Modell beim indirekten Leseweg beschrieben ist (Mayer, 2013). Voraussetzung für diese Strategie ist die Einsicht in die lautliche Struktur der Sprache und die Kenntnis der Buchstaben (Marx, 2007b). In dieser Phase setzt ein entscheidender Strategiewechsel ein: Die Kinder entdecken, dass Grapheme einen Lautwert (Phonem) symbolisieren (Niedermann & Sassenroth, 2002). Neue Wörter und Pseudowörter können mit dieser Strategie phonologisch rekodiert werden. Die Entwicklung des phonologischen Rekodierens hängt zum einen mit der Buchstabenkenntnis und zum anderen mit der phonologischen Bewusstheit zusammen (Mayer, 2016; Richter & Müller, 2017). Aufgrund der relativ konsistenten Graphemen-Phonemen-Zuordnung in der deutschen Schriftsprache kommt die Lautfolge der korrekten Aussprache in der Regel recht nahe – dies etwa im Gegensatz zum Englischen oder Französischen mit deutlich intransparenteren Orthographien (Mayer, 2016; Richter & Müller, 2017).

Kennzeichnend für die alphabetische Strategie ist die kognitiv herausfordernde Operation der Phonemsynthese (Zusammenschleifen von Lauten zu einem Wort) (Scheerer-Neumann, 2015). Die Herausforderung besteht u.a. darin, dass die phonologische Wortform nicht durch Addition von Lauten gebildet wird, sondern dass in einem integrativen Akt etwas qualitativ Neues entsteht (ebd.). Es wird gewissermaßen eine phonologische Rohform (ES:E:L) eines Wortes gebildet, das anschließend lexikalisch korrigiert werden muss, indem es mit einem Eintrag im mentalen Lexikon verglichen wird (Mayer, 2013, S. 29). Die Sinnentnahme erfolgt dadurch oft verzögert. Zusätzlich erschwerend kommt dazu, dass manche Kinder ihr phonologisches Arbeitsgedächtnis dadurch entlasten, dass sie wiederholend „(SS So:So:So::ne: Sonn(e)“ (Niedermann & Sassenroth, 2002, S. 10) oder gedehnt lesen „(rrrooosssaaa“ (Füssenich & Löffler, 2008, S. 83). Besondere Syntheseschwierigkeiten bieten längere Wörter und Wörter mit Konsonantenclustern (z.B. *Obstsalat*) (Scheerer-Neumann, 2015, S. 75), aber auch Buchstabenverbindungen wie <ie> oder das Dehnungs-<h> (Helbig et al., 2005, S. 31).

Hinsichtlich der Entwicklung im deutschsprachigen Raum zeigt sich mit Bezug auf die Studie von Klicpera und Gasteiger-Klicpera (1993), dass die meisten Kinder schon nach neun Wochen Unterricht viele unbekannte Wörter und Pseudowörter lesen – ein Hinweis auf die Anwendung der alphabetischen Strategie (Marx, 2007b). Aus dem Leselehrgang bekannte Wörter wurden zu diesem Zeitpunkt allerdings deutlich sicherer und schneller gelesen, was darauf schließen

lässt, dass bei bekannten Wörtern zumindest am Anfang noch Anteile der logographemischen Strategie angewandt wurden.

In der Regel werden Wörter im Laufe des zweiten Schulhalbjahres (erste Klasse) sehr konsequent und vollständig *erlesen* (phonologisch rekodiert), selbst dann, wenn Wörter in einem Text wiederholt vorkommen (Scheerer-Neumann, 2015) oder wenn Wörter bereits bekannt sind, z.B. der eigene Name (Niedermann & Sassenroth, 2002). Kinder scheinen in dieser Anfangsphase an diesem sehr sicheren, aber ineffizienten indirekten Leseweg festzuhalten, was sich aus empirischen Studien (vgl. Klicpera & Gasteiger-Klicpera, 1993) ableiten lässt, die zeigen, dass die Kinder bei bekannten Wörtern bis etwa Mitte des ersten Schuljahres ihre Lesegeschwindigkeit kaum steigern (Marx, 2007b). Außerdem zeigt sich bei Kindern der ersten Klasse eine sehr hohe Korrelation ($r = .91$) in der Lesegeschwindigkeit für echte Wörter und Pseudowörter, ein weiterer Beleg für die Anwendung der phonologischen De- bzw. Rekodierstrategie (Landerl & Wimmer, 2008, S. 154). Im Gegensatz zur logographemischen Phase wird in diesem Entwicklungsabschnitt der bildliche Kontext als Entschlüsselungshilfe regelrecht gemieden. Die Kinder richten ihre ganze Aufmerksamkeit darauf, allen Schriftzeichen die entsprechenden Laute zuzuordnen (Niedermann & Sassenroth, 2002).

Die alphabetische Strategie entwickelt sich in verschiedenen Zwischenstadien von einer *beginnenden* über eine *teilweise entfalteten* zu einer *weitgehend entfalteten* und schließlich zur *voll entfalteten alphabetischen* Strategie (Helbig et al., 2005; Niedermann & Sassenroth, 2002). Gegen Ende der ersten Klasse scheint die Worterkennung schon weitestgehend automatisiert zu sein, und Kinder mit unauffälliger Leseentwicklung verwenden zunehmend Anteile der orthographischen Strategie (Füssenich & Löffler, 2008; Marx, 2007b). Die voll entfaltete alphabetische bzw. beginnende orthographische Strategie geht mit einem Zuwachs an Lesegeschwindigkeit einher. Dies hängt damit zusammen, dass funktionale Einheiten zunehmend gebündelt (häufig vorkommende Buchstabencluster bzw. Signalgruppen, Silben, Morpheme) und dass die beiden Prozesse (indirekter und direkter Leseweg) stärker in Interaktion treten und automatisiert werden (Füssenich & Löffler, 2008; Klicpera et al., 2013). Insbesondere ist die Nutzung von Silbenstrukturen von besonderer Bedeutung, da die kognitiv aufwendige Zwischenspeicherung von Phonemen und die nachfolgende Synthese überflüssig werden. Das Lesen wird dadurch weniger fehleranfällig und effizienter, wie Richter und Müller (2017) auf der Grundlage empirischer Studien feststellen.

Orthographische Strategie: Automatisierung auf sublexikalischer und lexikalischer Ebene

Diese Phase kann auch als „Fortgeschrittenes Lesen“ (Niedermann & Sassenroth, 2002, S. 11) bezeichnet werden, das sich laufend aus den bereits beschriebenen Prozessen einer gefestigten alphabetischen Strategie entwickelt (Richter & Müller, 2017). Im Modell von Klicpera et al. (2017, S. 29) wird diese Phase auch als „partiell lexikalisches Lesen“ bezeichnet, das sich bei unauffällig lesenden Kindern etwa ab der zweiten Klasse entwickelt und sukzessive in das „lexikalische Lesen“ übergeht.

Auf Basis orthographischer Repräsentationen werden sublexikalische Wortteile (Morpheme, Silben, Signalgruppen) oder Wörter („Sichtwortschatz“) nun direkt erkannt, und das flüssige Lesen dadurch begünstigt (Richter & Müller, 2017, S. 58; Scheerer-Neumann, 2018, S. 78). Durch die ganzheitliche Verarbeitung größerer orthographischer Einheiten der Schriftsprache ist ein wesentlich effizienteres Lesen möglich und das Arbeitsgedächtnis wird für weitere inhaltliche Verarbeitungsprozesse (Sinnentnahme) frei (Mayer, 2013). Durch das Segmentieren in größere sublexikalische Verarbeitungseinheiten kommt es zu einer Vergrößerung der im Kurzzeitgedächtnis gespeicherten Einheiten bzw. zur Reduzierung der Anzahl der Schritte, die beim Erlesen notwendig sind (Scheerer-Neumann, 2018). Die Anzahl speicherbarer Items bei Kindern wird dabei auf drei bis sieben geschätzt (Niedermann & Sassenroth, 2002, S. 12; Wember, 1999, S. 20). Die Assoziationen zwischen sublexikalischen Einheiten und deren Phonologie tragen zur Automatisierung bei, da sie nicht an einzelne Wörter gebunden sind, sondern der Erkennung zahlreicher Wörter dienen (z.B. <iege> in Ziege, Liege, siegen) (Mayer, 2013, S. 33). Trotz dieses direkten, über das orthographische Lexikon gelenkten Lesewegs unterscheidet sich diese Phase aber deutlich vom logographemischen Lesen, da nicht mehr visuelle Merkmale, sondern phonologisch strukturierte graphemische Repräsentationen von Wortteilen und Wörtern genutzt werden (Günther, 2007; Scheerer-Neumann, 2018). Das heißt, dass auch in dieser Phase phonologische Rekodierprozesse weiterhin bedeutsam bleiben, insbesondere wenn sie die Nutzung von Silbenstrukturen einschließen, da neue und seltene Wörter auch von geübten Leserinnen und Lesern auf diesem Weg erschlossen werden müssen (Richter & Müller, 2017).

In der Entwicklung des Leseprozesses stellt die orthographische Strategie, d.h. die Automatisierung, den entscheidenden Schritt auf dem Weg zum kompetenten Leser bzw. zur kompetenten Leserin dar (Mayer, 2016). Durch die zunehmende Verwendung der orthographischen Strategie auf Wortebene nimmt die Lesegeschwindigkeit vom ersten zum zweiten Schuljahr deutlich zu. Gemäß der Studie von Klicpera und Gasteiger-Klicpera (1993) liegt sie am Ende des ersten Schuljahres im Durchschnitt bei ca. 30 Wörtern und am Ende des zweiten Schuljahres

bei 50 Wörtern pro Minute. Der Übergang zur orthographischen Strategie und deren Automatisierung stellt für leseschwache Kinder denn auch eine entscheidende Hürde bei der Entwicklung basaler Lesefertigkeiten dar (Mayer, 2013).

Integrativ-automatisierte Strategie: Weiterentwicklung der visuellen Worterkennung

In dieser letzten Phase basaler Lesefertigkeiten auf Wortebene erwirbt das Kind keine neuen Fertigkeiten, sondern automatisiert die bereits erworbenen Fähigkeiten, weshalb diese Phase denn auch den Übergang zum flüssigen Lesen kennzeichnet (Niedermann & Sassenroth, 2002). Die Zuverlässigkeit der phonologischen Rekodierung und direkten Worterkennung verbessert sich zu Beginn der Grundschulzeit besonders stark, kann sich aber im weiteren Verlauf der Grundschulzeit noch steigern (Richter & Müller, 2017). Gemäß dem Forschungsüberblick von Richter und Müller (2017) stellt sich beim phonologischen Rekodieren einfacher Pseudowörter und beim Erkennen einfacher Wörter aus dem Grundwortschatz spätestens Ende der vierten Klasse ein Deckeneffekt ein, was darauf hindeutet, dass basale Lesekompetenzen von der überwiegenden Zahl von Kindern prinzipiell beherrscht werden. Die Automatisierung der Worterkennung entlastet den Arbeitsspeicher und die Aufmerksamkeit zugunsten der Prozesse beim Textverstehen (Scheerer-Neumann, 2018). Hierarchiehöhere Prozesse (Kap. 2.3.3) gewinnen folglich zunehmend an Bedeutung, auch wenn die Entwicklung des Sichtwortschatzes und die Routinisierung der Worterkennung – anders als die Zuverlässigkeit von Worterkennungsprozessen – am Ende der Grundschulzeit auch bei Kindern mit normaler Leseentwicklung nicht abgeschlossen ist (Richter & Müller, 2017).

2.7 Textschwierigkeit und Leseaufgaben

Passung zwischen Lesefähigkeit und Textschwierigkeit

In den bisherigen Ausführungen wurde vor allem die Seite der lesenden Person betrachtet. Lesen und Leseverständnis sind aber in hohem Masse von den Verstehensanforderungen und den Aufgaben, die durch den Text gestellt werden, beeinflusst (Artelt, 2009). Geht es etwa um das Heraussuchen isolierter Informationen, die prominent platziert sind, oder um hierarchiehöhere Verstehensanforderungen? (Artelt, 2009; Lindauer & Schneider, 2011).

Rosebrock et al. (2011) vermuten, dass die Vereinfachung von Texten im Unterrichtsalltag eher selten vorkommt. Es sei aber dennoch sinnvoll, über ein solches Handlungswissen zu verfügen. Auch Philipp (2012a) stellt fest, dass es für Schülerinnen und Schüler oft schwierig ist, Texte (etwa im Fachunterricht) zu verstehen und dass die Lehrpersonen deshalb gefordert sind, eine

Brücke zwischen dem Text und den Lesenden zu schlagen. Die Brücke kann beispielsweise mit dem Vermitteln von Lesestrategien oder der Auswahl geeigneter Texte geschlagen werden. Dafür ist ein Basiswissen über pädagogisch relevante Textmerkmale zentral (Wember, 1999). Modelle der Vorhersage von Textschwierigkeiten unterscheiden sich darin, auf welcher Ebene „Textschwierigkeit“ definiert wird (Artelt, 2009, S. 131). Die Schwierigkeit kann z.B. an einfachen Kriterien nach dem Prinzip „Kurze Sätze, vertraute Wörter“ bis hin zu relativ komplexen linguistischen Merkmalen wie grammatikalische Komplexität und Menge redundanter Informationen festgemacht werden (ebd., S. 131). Zur Beschreibung der textbezogenen Eigenschaften, welche die Leseflüssigkeit und das Leseverständnis entscheidend beeinflussen, wird die „Leserlichkeit“ und die „Lesbarkeit“ unterschieden (Herné & Löffler, 2014, S. 147). Die Leserlichkeit betrifft die drucktechnische Gestaltung wie z.B. Schriftgröße, Schriftart und Absatzformatierungen (ebd.). Die Lesbarkeit betrifft dagegen die sprachliche Einfachheit im Sinne von Wortlänge, Wortschatz, Satzlänge, Satzverknüpfungen (Lindauer & Schneider, 2011; Rosebrock et al., 2011). Weitere Ebenen zur Bestimmung der Textschwierigkeit betreffen die kognitive Gliederung eines Textes sowie inhaltliche und motivationale Aspekte (Rosebrock et al., 2011). Mit Verweis auf entsprechende Studien (Christmann & Groeben, 1999) wird die Bedeutung der kognitiven Gliederung (z.B. Vorstrukturierung durch globale Angaben, was im Text folgt / Wiederholungen und Zusammenfassungen / Hervorhebungen, Überschriften) und der Lesbarkeit als wichtigste Dimension bezeichnet (Rosebrock et al., 2011, S. 66). Die hier interessierende Ebene betrifft die Lesbarkeit von Texten durch deren lesetechnische Anforderungen, da sie insbesondere im Zusammenhang mit basalen Lesekomponenten (Leseflüssigkeit, Satzverstehen) entscheidend ist (ebd.).

Oberflächenstruktur und Lesbarkeitsindex

Für die Förderung basaler Lesekompetenzen, die sich zunächst auf Wort- und Satzebene abspielt, ist eine Passung der sprachlichen Oberflächenkomplexität eines Textes an die Kompetenzen auf Leserseite entscheidend (Rosebrock et al., 2011). Um die Textschwierigkeit auf einer objektiveren Basis beurteilen zu können, wurden in der Leseforschung verschiedene Lesbarkeitsformeln entwickelt, die an der sprachlichen Oberfläche eines Textes ansetzen und etwa Wort- oder Satzlänge, Silbenstruktur, Wörter pro Satz usw. berücksichtigen (Lindauer & Schneider, 2011). Textschwierigkeiten können hier einfach ermittelt und in Zahlwerten ausgedrückt werden.

Neben zahlreichen relativ aufwändigen Kennwerten zur Beurteilung von Texten sind für die Praxis einfache und schnell anwendbare, aber trotzdem informative Verfahren – wie etwa der

Lesbarkeitsindex (Lix) oder Flesch-Index interessant (Wember, 1999). Insbesondere hat sich der von Björnsson (1968) entwickelte Lix, der zwar nur wenige Oberflächenmerkmale eines Textes berücksichtigt, aufgrund seiner hohen Inhaltsvalidität und der praktischen Anwendbarkeit als geeignet erwiesen und wird in zahlreichen Fachpublikationen kritisch gewürdigt (Herné & Löffler, 2014; Rosebrock et al., 2011; Scheerer-Neumann, 2018; Wember, 1999). Der Lix (siehe Formel, Abbildung 5) basiert auf lediglich zwei Textmerkmalen – der durchschnittlichen Satzlänge und dem Prozentsatz der Wörter mit mehr als sechs Buchstaben (Wember, 1999).

$$\text{Lix} = \frac{\text{Zahl der Wörter}}{\text{Zahl der Sätze}} + 100 * \frac{\text{Zahl der Wörter mit mehr als sechs Buchstaben}}{\text{Zahl der Wörter}}$$

Abbildung 5. Formel für die Berechnung des Lix (Wild & Pissarek, 2018, S. 9)

Diese beiden Kennwerte (Satzlänge und Wortlänge) konnten neben der Häufigkeit von Wortwiederholungen, für die Schwierigkeitsbestimmung in der Leseforschung immer wieder bestätigt werden (Wember, 1999). Erklären lässt sich dies etwa damit, dass für die Wortebene gilt: je kürzer ein Wort, umso verständlicher ist es in der Regel auch. Statistisch gesehen, sind häufig gebrauchte Wörter kürzer als seltenere Wörter. Auf Satzebene gilt entsprechend, dass kürzere Sätze meist auch besser verständlich sind (Rosebrock et al., 2011). Insbesondere haben Forschungen (z.B. Bamberger & Vanecek, 1984) gezeigt, dass die nach Lix eingestuft Texten für Lesende unterschiedlichen Kompetenzniveaus tatsächlich verschieden anspruchsvoll waren, und die Autoren bescheinigen dem Lix bei hoher Anwendungsökonomie eine vergleichbar hohe Validität wie den vergleichsweise komplexeren Indices (Rosebrock et al., 2011). Dennoch ist der Lix lediglich als grobes Schätzverfahren zu betrachten (Wember, 1999), da zahlreiche weitere Faktoren die Schwierigkeit eines Textes bestimmen, was im Folgenden erläutert wird.

Kriterien der Textvereinfachung

Wie bereits angesprochen, lässt sich die Schwierigkeit eines Textes streng genommen nur mit Blick auf die Kompetenzen der lesenden Person definieren. Außerdem spielen neben den erwähnten Merkmalen auf Textoberfläche auch Elemente der Tiefenstruktur (Bedeutungsnetz) eine zentrale Rolle (Kernen & Riss, 2012). Hier sei auf das *Hamburger Verständlichkeitskonzept* (Langer, Schulz von Thun & Tausch, 2011) verwiesen, bei dem solche Tiefenstrukturen in einer fünfstufigen Skala bestimmt werden. Im Folgenden soll der Blick aber auf Maßnahmen zur Textvereinfachung gerichtet werden, die enger an die Erwerbsphase basaler Lesekompetenzen gebunden sind und mehrheitlich die Oberflächenstruktur betreffen.

Für die *Wortauswahl* haben Forster und Martschinke (2003, S. 61) einige Kriterien aufgestellt:

- 1) Kurze Wörter mit einfacher Konsonanten-Vokal-Struktur sind einfacher zu lesen als Wörter mit Konsonantenclustern, die zudem lang sind (mehr als sechs Buchstaben).
- 2) Lange Vokale sind einfacher zu analysieren als kurze (Rabe ist leichter als Dach) und Dauerkonsonanten (l, m, s) sind einfacher zu analysieren als Plosive (t, p, k). Hierzu lässt sich mit Bezug auf empirisch und theoretisch gesicherte Ergebnisse (Groeben & Christmann, 1996) ergänzen, dass Sätze aus Wörtern mit wenigen Silben, die zudem statistisch häufig vorkommen und dem Lesenden vertraut sind, leichter gelesen werden (Rosebrock et al., 2011).
- 3) Anschauliche Wörter sind für Kinder leichter verständlich als Abstrakta (= Wörter die oft auf -heit, -keit, -ung, -ion usw. enden) (ebd.).
- 4) Auf Satzebene ist die Auflösung von hypotaktischen Satzgefügen zentral. Das heißt, dass Sätze mit untergeordneten Nebensätzen zugunsten kurzer gleichartiger Sätze, die vorzugsweise im Aktiv statt im Passiv formuliert sind, aufgelöst werden (Mayer, 2015; Rosebrock et al., 2011).
- 5) Der Einsatz von Kohäsionsmitteln hat einen Einfluss auf die Verständlichkeit von Sätzen und Texten. „Unter Kohäsionsmitteln werden in der Linguistik sprachliche Mittel verstanden, mit denen Beziehungen zwischen Wörtern, Satzteilen und Sätzen ausgedrückt werden“ (Mayer, 2015, S. 225). Substitutionen (Textelemente, die wieder aufgenommen, aber durch ein anderes Wort ersetzt wird) und nicht eindeutig zuzuordnende Pronomen (z.B. Personalpronomen, Demonstrativpronomen) erhöhen die Komplexität. Beispiel: *Der Mann kämpfte tapfer mit dem Bär. Er war jedoch schwächer und floh.* Optimierte Version: *Der tapfere Mann kämpfte mit dem Bär. Der Mann war aber schwächer und rannte weg.*
- 6) Weitere textvereinfachende Merkmale beziehen sich auf die Chronologie, in der Informationen gegeben werden, auf die Zeitform (Präsens) und die Informationsdichte bzw. den Verzicht auf unwesentliche Informationen (ebd.).

Kindern soll nicht vorenthalten sein, ihre schriftsprachlichen Kompetenzen durch anspruchsvolle Texte zu erweitern. Es geht aber darum, Texte zu ermitteln, die dem individuellen Instruktions- oder Unabhängigkeitsniveau entsprechen (siehe Niveaus der Leseflüssigkeit: Rosebrock et al., 2011). Dies ist nicht nur aus motivationalen Gründen wichtig, sondern auch deshalb, weil Kinder aus schwierigen Texten schlechter lernen und Erfolge bei Lautleseverfahren größer sind, wenn die Textschwierigkeit den Lesefähigkeiten der Kinder angepasst ist (ebd.).

2.8 Merkmale der Leseschwierigkeit

Kinder mit Leseschwierigkeiten durchlaufen in ihrer Leseentwicklung grundsätzlich dieselben Prozesse wie unauffällige Kinder: Die Unterschiede zeigen sich vor allem in einer zeitlichen Verschiebung und dem größeren Aufwand zum Erwerb von Lesekompetenzen (Fischer & Gasteiger-Klicpera, 2013; Hartmann, 2007; Scheerer-Neumann, 2018; Warnke, Hemminger & Plume, 2004). Damit geht einher, dass Kinder mit Lese-Rechtschreibschwierigkeiten grundsätzlich *keine anderen*, sondern *viel mehr* Fehler machen als gleichaltrige Kinder mit unauffälliger Leseentwicklung (Herné & Löffler, 2014, S. 66). Dies bedeutet nicht, dass sich Leseschwierigkeiten irgendwann von selbst auswachsen und betroffene Kinder dieselben Lesekompetenzen erreichen wie andere Kinder bzw. Jugendliche oder Erwachsene (Warnke et al., 2004). Im Gegenteil: Kinder mit Lese-Rechtschreibschwierigkeiten brauchen gezielte Unterstützung beim Erwerb der verschiedenen Teilprozesse. Ganz offensichtlich stagnieren Kinder und Jugendliche ohne diese Unterstützung auf einem Leseniveau, das es ihnen verunmöglicht, mehrere Informationen aufeinander zu beziehen und folgerichtige Schlüsse zu ziehen, wie die „Adult Literacy and Lifeskills Survey“ (Sturm, 2012) zeigt.

Im Folgenden werden zentrale Schwierigkeiten leseschwacher Kinder beschrieben, die auf kognitionspsychologischen Modellen beruhen und sich sowohl auf hierarchieniedrige als auch auf hierarchiehohe Prozesse beziehen. Außerdem sollen mögliche Defizite in den Vorläuferfertigkeiten kurz angesprochen werden.

2.8.1 Schwierigkeiten mit der Phonologischen Informationsverarbeitung

Defizite leseschwacher Kinder zeigen sich oft schon vor dem eigentlichen Beginn des Schriftspracherwerbs. Im Zusammenhang mit den besprochenen Vorläuferfertigkeiten (Kap. 2.5) wird in diesem Zusammenhang die sogenannte „Double-Deficit Hypothese“ (Wolf & Bowers, 1999) diskutiert. Sie besagt, dass Kinder mit Lese-Rechtschreibschwierigkeiten bereits früh durch Defizite in der phonologischen Bewusstheit *und* Benennungsgeschwindigkeit auffallen (Klicpera et al., 2017; Mayer, 2016). Kinder können aber auch unabhängig von Schwierigkeiten in der phonologischen Bewusstheit Probleme in der Benennungsgeschwindigkeit haben, und umgekehrt können Probleme in der phonologischen Bewusstheit ohne Defizite bei der Benennungsgeschwindigkeit auftreten (Klicpera et al., 2017).

Gut belegt ist, dass sich leseschwache Kinder vor allem durch Defizite in der Benennungsgeschwindigkeit von später unauffällig lesenden Kindern unterscheiden, wobei in einer späteren Phase (Ende erste und zweite Klasse) insbesondere Auswirkungen auf die Lesegeschwindigkeit zu vermerken sind (Klicpera et al., 2017; Marx, 2007b; Mayer, 2018a; Wolf & Bowers, 1999).

Mit Verweis auf eine Vielzahl empirischer Studien – sowohl aus dem englischsprachigen wie auch dem deutschsprachigen Raum – kommt Mayer (2016) insgesamt zum Schluss, dass Kinder mit Lese-Rechtschreibschwierigkeiten meist wegen Defiziten in der phonologischen Bewusstheit wie auch der Benennungsgeschwindigkeit zu erkennen sind. Die Leseleistungen (Lese-geschwindigkeit und Leseverständnis) von Kindern mit einem doppelten Defizit liegen am Ende der ersten und zweiten Klasse deutlich unter dem Mittelwert der Kinder ohne Defizite (Mayer, 2014). Auch Defizite in nur einem Bereich wirken sich gemäß der Studie auf die Leseleistung aus, jedoch in etwas geringerem Masse als beim doppelten Defizit. Auch Wimmer (1993) konnte in seiner Studie nachweisen, dass die Benennungsgeschwindigkeit der beste Prädiktor war, um zwischen leseschwachen und durchschnittlich lesenden Kindern zu differenzieren. Defizite in der phonologischen Bewusstheit scheinen hingegen im Zusammenhang mit Leseschwierigkeiten weniger konsistent, da Kinder im Verlaufe des Anfangsunterrichts (bis Ende zweite Klasse) Defizite in diesem Bereich eher überwinden können – zumindest im deutschsprachigen Raum mit einer relativ transparenten Orthographie (Mayer, 2016; Wimmer, 1993). In diesem Sinne sind Defizite in der phonologischen Bewusstheit kein zwingendes Merkmal leseschwacher Kinder (Gorecki & Landerl, 2015).

Schließlich zeigt der Blick auf empirische Ergebnisse zur dritten Komponente der phonologischen Informationsverarbeitung Zusammenhänge zwischen dem phonologischen Arbeitsgedächtnis und der Worterkennung (Marx, 2007b; Schneider, 2017). Kinder mit Defiziten in der Worterkennung haben gemäß Studien Schwierigkeiten mit dem Nachsprechen von Pseudowörtern („phonologischer Buffer“), nicht aber mit dem Nachsprechen von Wort- und Zahlenfolgen („phonologisches Rehearsal“) (Mayer, 2013, S. 43). Weiter zeigt eine Langzeitstudie (Seigneuric & Ehrlich, 2005) mit Kindern der ersten bis dritten Klasse, dass sich Defizite im Arbeitsspeicher insbesondere auf hierarchiehöherer Ebene, also beim Leseverständnis bemerkbar machen.

2.8.2 Schwierigkeiten beim basalen Lesen

Im Bereich hierarchieniedriger basaler Lesekompetenzen zeigen leseschwache Kinder eine Reihe von Defiziten, welche die Aneignung der Graphem-Phonem-Korrespondenz, das phonologische Rekodieren, das Synthetisieren sowie die direkte Worterkennung betreffen. Die Leseflüssigkeit, die die Verbindung zwischen Wortlesen und Textverständnis darstellt (Scheerer-Neumann, 2018), wird aufgrund ihrer herausragenden Bedeutung im nachfolgenden Kap. 2.8.3 gesondert beschrieben.

Graphem-Phonem-Korrespondenz

Erste Anzeichen für eine Leseschwierigkeit lassen sich bereits ca. drei Monate nach Schulbeginn daran erkennen, dass sich Kinder im Aufbau der Graphem-Phonem-Korrespondenz schwer tun und in der Regel deutlich geringere Buchstabenkenntnisse besitzen als ihre Peers (Günther, 2007; Klicpera et al., 2017; Schneider, 2017). Auch wenn sich die Buchstabenkenntnis nur begrenzt zur Identifikation leseschwacher Kinder eignet, so scheint sie gemäß Längsschnittstudien doch in einem Zusammenhang mit späteren Lesekompetenzen zu stehen (Marx, 2007b; Schneider, 2017). Leseschwache Kinder zeigen bereits in der beginnenden alphabetischen Phase einen deutlichen Rückstand im Vergleich zu unauffällig lesenden Kindern (Klicpera et al., 2017). Bereits nach drei Monaten konnten leseschwache Kinder dieser Studie im Gegensatz zu gut lesenden Kindern die bis dahin gelernten Buchstaben (ca. acht) nicht richtig benennen und nur zwei Drittel der vertrauten und häufig gelesenen Wörtern fehlerfrei erkennen (ebd., S. 153). Dies verweist auf die erschwerten Behaltensleistungen der Graphem-Phonem-Korrespondenz. Zu ähnlichen Ergebnissen kommen auch Wimmer und Hummer (1990) in ihrer Studie mit Kindern in der zweiten Hälfte der ersten Klasse.

Die zu Beginn des Leseerwerbs häufig beobachtete Verwechslung visuell ähnlicher Grapheme (z.B. , <d> / <m>, <w>) oder phonologisch ähnlicher Phoneme (z.B. /b/, /p/,) ist dagegen ein eher allgemeines Phänomen und zeigt sich auch bei nicht-leseschwachen Kindern (Herné & Löffler, 2014, S. 66; Scheerer-Neumann, 2018, S. 79). Im Zusammenhang mit dem Aufbau der Phonem-Graphem-Korrespondenz werden oft auch Defizite in der visuellen Wahrnehmung bzw. der optischen Differenzierungsfähigkeit von Buchstaben angenommen. Für diese These finden sich allerdings keine Belege. Die visuellen Wahrnehmungsleistungen stehen in einem geringen Zusammenhang mit späteren Leseleistungen und tragen kaum zur Erklärung von Unterschieden zwischen guten und schlechten Leserinnen und Lesern bei (Marx, 2007b; Richter & Christmann, 2002; Tacke, 2011).

Phonologisches Rekodieren und Phonemsynthese

Mit dem Verständnis der Graphem-Phonem-Korrespondenz geht die Ausbildung des phonologischen Rekodierens einher (Fischer & Gasteiger-Klicpera, 2013). Phonologische Rekodierungsprozesse scheinen gemäß Übersichtsstudie von Richter und Christmann (2002) „eine potenzielle Quelle von Lesefähigkeitsdifferenzen“ (S. 38) darzustellen. Leseschwachen Kindern gelingt zwar das Lautieren von Buchstaben, aber nicht das Zusammenlauten bzw. Synthetisieren (Phonemsynthese) von Buchstabenfolgen (Klicpera et al., 2017; Schröder-Lenzen, 2009). Sie sind dadurch oft nicht in der Lage, die Aussprache eines Wortes zu erschließen (Fischer &

Gasteiger-Klicpera, 2013; Tacke, 2011). Zudem scheinen leseschwache Kinder aufgrund des mühsamen Erlesens im Sinne einer Ausweichstrategie auch immer wieder auf die logographische Lesestrategie zurückzufallen und bleiben übermäßig lange auf der logographischen Stufe stehen (Fischer & Gasteiger-Klicpera, 2013; Scheerer-Neumann, 2018; Tacke, 2008).

Die phonologische Rekodierung von Wörtern scheint ein persistierendes Merkmal leseschwacher Kinder zu sein (Schründer-Lenzen, 2009). Besonders augenfällig werden die Schwierigkeiten beim phonologischen Rekodieren von Pseudowörtern oder neuen Wörtern (Schründer-Lenzen, 2013). Gemäß Studie von Klicpera, Graeven und Schabmann (2000) können leseschwache Kinder einige Monate nach Schulbeginn deutlich weniger Pseudo- und unbekannte Wörter richtig rekodieren als durchschnittlich lesende Kinder, und auch das Zusammenschleifen gelingt ihnen deutlich schlechter. Leseschwache Kinder lasen zu diesem Zeitpunkt auch kurze Pseudowörtern (drei bis fünf Buchstaben) in weniger als einem Fünftel der Fälle richtig (Klicpera et al., 2017, S. 153). Ein praktischer Richtwert für eine erste informelle Diagnose stellt die „drei Monatsmarke“ dar:

Leseanfänger, die innerhalb der ersten zwei bis drei Monate Leseunterricht die Lesesynthese verstehen, nehmen in der Regel eine problemlose weitere Leseentwicklung, wohingegen Kinder, die auch Ende des ersten Schuljahres unbekannte Wörter noch nicht ohne Hilfestellungen erlesen können, nachhaltig in ihrer weiteren Leseentwicklung gefährdet sind. (Schründer-Lenzen, 2013, S. 47)

Mit fortschreitender alphabetischer Strategie zeigt sich ein weiteres Problem leseschwacher Kinder darin, dass sie orthographische Einheiten wie Silben, Morpheme und Signalgruppen nicht zusammenfassen (Ritter, 2005; Scheerer-Neumann, 1981; Tacke, 2011). Gute Leserinnen und Leser hingegen segmentieren Wörter intuitiv in Silben und Buchstabengruppen (Tacke, 2011).

Neben diesen lesetechnischen Schwierigkeiten scheint auch das Korrekturverhalten leseschwacher Kinder oft ungünstig bzw. ineffektiv. Entweder gelingt es ihnen gar nicht, Fehler überhaupt zu erkennen (Fischer & Gasteiger-Klicpera, 2013; Warnke, Hemminger, Roth & Schneck, 2002), oder die Korrektur führt immer weiter weg vom Zielwort, so dass völlig unsinnige Wortbildungen entstehen (Schründer-Lenzen, 2009). Hilfestellungen der Lehrperson werden in der Regel nicht konstruktiv aufgegriffen, und die Kinder verharren in ihrer ungünstigen Strategie (ebd.). In dieser Phase des Leseerwerbs zeigt sich zudem oft ein weiteres Problem: Kinder versuchen, ihre Defizite durch Ratestrategien zu kompensieren, indem sie den Kontext wie z.B. Bilder aus Leselehrgängen nutzen (Moll & Landerl, 2010). Kinder, die solche Strategien nutzen, lesen bei längeren Wörtern nur den Wortanfang und erraten den Rest des Wortes – eine

Strategie, die längerfristig nicht erfolgreich ist (Moll & Landerl, 2010; Scheerer-Neumann, 2018).

Erwerb der orthographischen Strategie und Aufbau des Sichtwortschatzes

Die Unterschiede zwischen schwach und gut lesenden Kindern werden meist etwa ab der zweiten Hälfte des ersten Schuljahres offensichtlich (Klicpera et al., 2017). Die Schere zwischen unauffällig und schwach lesenden Kindern öffnet sich immer weiter, wie die Metaanalyse von Pfost, Hattie, Dörfler und Artelt (2014) zeigt. Beim Erwerb der orthographischen Strategie (automatisierte Worterkennung) zeigen sich die bereits angesprochenen Schwierigkeiten des Nutzens größerer Einheiten, wodurch sich der Aufbau eines Sichtwortschatzes bei leseschwachen Kindern verzögert (Scheerer-Neumann, 2018). Unter Sichtwörtern werden Wörter zusammengefasst, „die im mentalen orthographischen Lexikon gespeichert sind und holistisch, das heißt ohne den Umweg über die Analyse und Synthese von einzelnen Buchstaben, erkannt werden können“ (Nix, 2011, S. 77f.).

Auf dieser Ebene des Wortlesens gelingt leseschwachen Kindern die Koordination zwischen Phonologie, Orthographie und Semantik nur unzureichend (Fischer & Gasteiger-Klicpera, 2013). Während gut lesende Kinder im Aufbau ihres inneren Lexikons auf orthographische Regularitäten achten, scheinen leseschwache Kinder diese Redundanzen bei Wörtern (wiederkehrende Silben, Morpheme, Signalgruppen) weniger zu nutzen, was den Leseprozess fehleranfälliger, vor allem aber auch langsamer macht (Schröder-Lenzen, 2013). Mit Blick auf verschiedene empirische Studien zeigt sich, dass leseschwache Kinder auch bei wiederholter Darbietung von Wörtern in der Worterkennungsgeschwindigkeit zurückbleiben (Klicpera et al., 2017). Leseschwache Kinder „erlesen“ am Ende der ersten Klasse selbst bekannte Wörter noch öfter, indem sie gedehnt lesen (53% der Wörter), wohingegen gut lesende Kinder zu diesem Zeitpunkt nur noch in 23% der Fälle auf diese indirekte Lesestrategie zurückgreifen und entsprechend in den meisten Fällen Wörter spontan erkennen (ebd., S. 156). Zur Überwindung dieser Art von Schwierigkeit sind Programme zur Vergrößerung von Worteinheiten und zum Aufbau des Sichtwortschatzes sinnvoll (Mayer, 2018b). Außerdem sind positive Effekte des wiederholten lauten Lesens auf die Worterkennung und Leseflüssigkeit empirisch gut belegt (Bertschi-Kaufmann & Kappeler, 2010).

Zusammenfassend lässt sich sagen, dass leseschwache Kinder trotz des deutlichen Fortschritts in der Sicherheit und Geläufigkeit des Lesens ihren Rückstand in der Regel nicht aufholen können, was sich ab der zweiten und dritten Klasse vor allem im Lesen von Texten zeigt (Klicpera et al., 2017). Die zunehmenden Unterschiede können durch die kumulierten Entwicklungs-

vorteile der Lesegenauigkeit und vor allem der Effizienz visueller Worterkennung – bzw. diesbezügliche Nachteile bei leseschwachen Kindern – erklärt werden. Dies wiederum wirkt sich in der Folge auf hierarchiehöhere Leseprozesse (Textverständnis) und weitere Ebenen des Lesens (Subjektebene und soziale Ebene) aus (Richter & Christmann, 2002; Richter & Müller, 2017).

2.8.3 Schwierigkeiten mit der Leseflüssigkeit

Die Leseflüssigkeit, d.h. das Konglomerat verschiedener Teilprozesse (Kap. 2.3.2), erweist sich im weiteren Verlaufe der Leseentwicklung als zentrales Unterscheidungskriterium zwischen gut und schwach lesenden Kindern. Das phonologische Rekodieren von Wörtern gelingt den meisten Kindern – auch leseschwachen – am Ende der ersten Klasse einigermaßen zuverlässig. Deutliche Unterschiede zeigen sich aber gemäß Langzeitstudien (Klicpera & Gasteiger-Klicpera, 1993; Landerl & Wimmer, 2008) in der *Dekodiergeschwindigkeit*. Demnach scheint die Lesegenauigkeit in konsistenten Orthographien mit fortschreitendem Lesen weniger gut zu differenzieren, die Worterkennungsgeschwindigkeit bleibt dagegen zentrales Unterscheidungskriterium (Mayer, 2016; Moll & Landerl, 2010; Walter, 2013). Die folgende Abbildung 6 zeigt die hohe Stabilität des Zurückbleibens lesewacher Kinder über die gesamte Schulzeit. Die Daten basieren auf der Wiener Längsschnittstudie von Klicpera und Gasteiger-Klicpera (1993, S. 54).

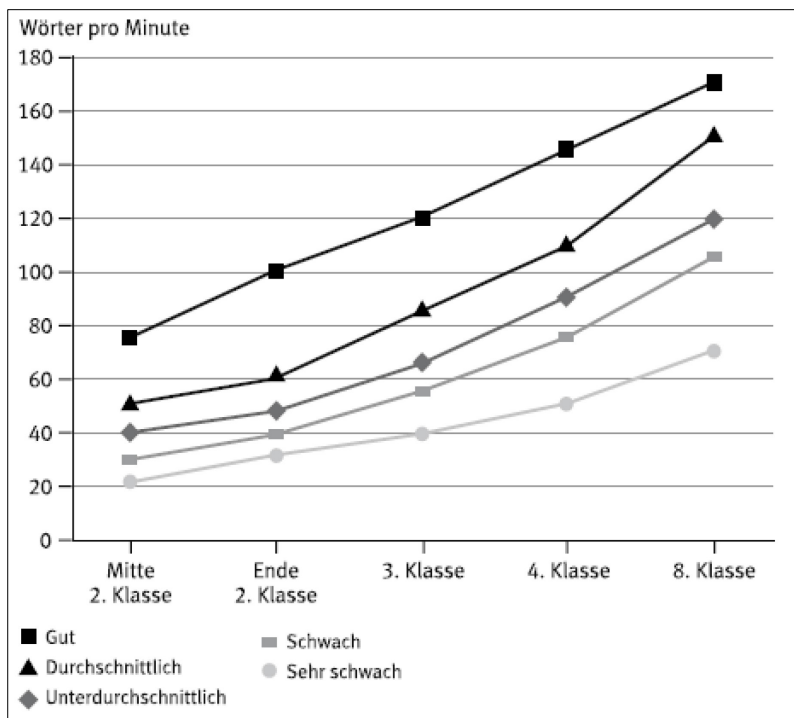


Abbildung 6. Entwicklung der Lesegeschwindigkeit (Wörter pro Minute) von der zweiten bis zur achten Klasse von Kindern mit anfänglich unterschiedlichen Lesefähigkeiten (entnommen aus Schneider, 2017, S. 83)

In Abbildung 6 wird deutlich, dass leseschwache Kinder ihren Rückstand, der bereits am Ende der ersten Klasse besteht, nicht aufholen können. Der Unterschied besteht selbst dann, wenn Kindern dieselben Wörter wiederholt dargeboten werden (Klicpera et al., 2017).

Auch der Blick auf Entwicklungsverläufe aus der Salzburger Längsschnittuntersuchung (erste bis achte Klasse) von Landerl und Wimmer (2008) zeigt, dass leseschwache Kinder über die gesamte Zeit zunehmend weniger durch fehlerhaftes Lesen als vielmehr durch eine äußerst stabile Differenz in der Lesegeschwindigkeit auffallen, sowohl bei bekannten Wörtern als insbesondere auch bei Pseudowörtern. In der Studie von Klicpera und Gasteiger-Klicpera (1993) wurde dagegen auch das fehlerhafte Lesen – insbesondere bei Pseudowörtern – als persistierendes Problem festgestellt. Den schwächsten Leserinnen und Lesern unterlaufen gemäß der Wiener Längsschnittstudie am Ende der vierten Klasse mehr Fehler als durchschnittlich Lesenden am Ende der ersten Klasse (Marx, 2007a).

Weiter zeigt sich, dass der Leseprozess nicht nur langsamer vor sich geht, sondern dass auch in der Prosodie Auffälligkeiten bestehen (Warnke et al., 2004). Die Intonation wirkt bei leseschwachen Kindern meist monoton. Es mangelt an angemessener Phrasierung des Gelesenen, wodurch zum einen das Verstehen des Lesenden, aber auch das Verständnis der Zuhörenden leidet (Niedermann & Hartmann, 2006). Allerdings ist dieses Problem, wie in Kap. 2.3.2 erwähnt, möglicherweise eher die Folge verlangsamter Automatisierungsprozesse. Zumindest ist ein korrelativer Zusammenhang zwischen gut entwickelten automatisierten Dekodierfähigkeiten und prosodisch korrektem Lesen empirisch belegt (Gold, 2009). Der Zusammenhang zwischen prosodischem Vorlesen und dem Leseverstehen beim stillen Lesen wird von Rasinski (2017, 520) mit Verweis auf entsprechende Forschungen aus dem englischsprachigen Raum dokumentiert. Dementsprechend können Defizite beim prosodischen lauten Lesen indirekte Hinweise für das Vorliegen einer Leseschwierigkeit geben (Scheerer-Neumann, 2018).

Zusammenfassend lässt sich die herausragende Bedeutung der Leseflüssigkeit nach Hartmann und Niedermann (2008) folgendermaßen beschreiben:

Wenn man geschriebene Wörter rasch und genau erkennt und Texte fließend lesen kann, stehen der Leserin bzw. dem Leser noch genügend kognitive Ressourcen zur Verfügung, die sie resp. er für höhere Verarbeitungsprozesse bzw. für das Textverstehen benötigt. Im Gegensatz dazu werden die kognitiven Energien beim unflüssigen Lesen hauptsächlich durch elementare (Wort-) Leseprozesse absorbiert, wodurch das Verstehen in der Regel erschwert ist. (S. 45)

Der Zusammenhang von Leseflüssigkeit und Leseverständnis ist in zahlreichen Studien belegt (u.a. Gold, 2009; Müller et al., 2013; Nix, 2011). Sie zeigen, dass Defizite in der Leseflüssigkeit mit Defiziten beim Leseverständnis assoziiert sind. Bevor eine Förderung auf hierarchiehöherer

Ebene wirksam sein kann, scheint es sinnvoll, zunächst die Leseflüssigkeit in den Blick zu nehmen. Rosebrock und Nix (2012) plädieren deshalb dafür, bei leseschwachen Kindern vorrangig auf die Leseflüssigkeit zu fokussieren und diese durch geeignete Lautlese-Verfahren (paired reading, Lesetandems, Lesetheater) zu trainieren, deren Wirksamkeit empirisch gut belegt ist (Gold, Nix, Rieckmann & Rosebrock, 2010; Kruse et al., 2010; Rosebrock et al., 2010; Schneider et al., 2013; Tacke, 2005; Therrien, 2004). Im Hinblick auf die Leseflüssigkeit und das Textverstehen haben sich insbesondere Lautlesetandems als erfolgreich erwiesen, bei denen Kinder mit unterschiedlicher Lesekompetenz (leseschwache und lesestarke Kinder) einander Texte synchron laut und wiederholend vorlesen (Rosebrock et al., 2010; Trenk-Hinterberger, Nix, Rieckmann, Rosebrock & Gold, 2007).

2.8.4 Schwierigkeiten mit dem Leseverständnis

Das Leseverständnis ist bei leseschwachen Kindern in besonderem Masse beeinträchtigt (Klicpera et al., 2017). Dies ist insofern fatal, da das eigentliche Ziel des Lesenlernens von Anfang an auf die Sinnentnahme aus Texten und das verständnisvolle Lesen gerichtet ist (ebd.). Der Zusammenhang mit Defiziten im basalen Lesen wurde bereits angesprochen und lässt sich vor dem Hintergrund von Informationsverarbeitungsmodellen damit erklären, dass das Dekodieren bei leseschwachen Kindern zu viel Aufmerksamkeit braucht und damit nicht genügend Speicher frei ist, das Gelesene auch zu verstehen (Günther, 2007; Wember, 1999). Insofern können Defizite im Leseverständnis die Folge vorangehender Defizite sein. Allerdings sind Beeinträchtigungen im basalen Lesen nicht alleine für Leseverständnisschwierigkeiten verantwortlich, da sich auch Hinweise darauf finden, dass Kinder trotz flüssigem Lesen deutliche Schwächen im Leseverständnis aufweisen können (Klicpera et al., 2017; Niedermann & Hartmann, 2006; Rosebrock & Nix, 2012; Scheerer-Neumann, 2018). Mit Bezug auf die Studie von Torppa et al. (2007) (erste und zweite Klasse) lassen sich verschiedene Profile identifizieren: 1) Kinder mit durchschnittlichen Dekodierleistungen einzelner Worte und deutlichen Schwierigkeiten im Leseverständnis 2) Kinder mit unterdurchschnittlichen Dekodierleistungen über die ersten beiden Schuljahre hinweg, aber ohne Leseverständnisschwierigkeiten 3) Kinder mit Defiziten in beiden Bereichen. „Diese Beispiele zeigen, dass von einer nennenswerten interindividuellen Variation von Entwicklungsverläufen auszugehen ist“ (Souvignier & Förster, 2011, S. 244). Auch gemäß Scheerer-Neumann (2018) lassen sich die genannten Profile leseschwacher Kinder identifizieren, wobei die Mehrheit der Kinder mit Leseverständnisproblemen *sowohl* bei hierarchieniedrigen Prozessen *als auch* auf der Ebene hierarchiehoher Prozesse (Satz- und Textebene) Defizite aufweisen. Neben den genannten Defiziten im basalen Lesen werden in verschiedenen Studien

übereinstimmend die folgenden Fehlleistungen diskutiert: 1) *Defizite, die mit dem Hörverstehen in Verbindung stehen*, 2) *Defizite im Bereich Wortschatz*, 3) *Defizite beim Bilden von Inferenzen auf Satz- und Textebene* sowie bei der *syntaktischen und semantischen Integration* und 4) *Defizite im Bereich der Anwendung von Lesestrategien* (Niedermann & Hartmann, 2006; Richter & Christmann, 2002; Scheerer-Neumann, 2018). Als weitere Komponente, die allerdings in der vorliegenden Arbeit nicht vertieft wird, scheinen sich Defizite im Arbeitsgedächtnis auf das Leseverständnis niederzuschlagen (Richter & Christmann, 2002; Seignuric & Ehrlich, 2005). Auch die Lesemotivation dürfte ein wichtiges unterscheidendes Merkmal zwischen lesestarken und leseschwachen Kindern sein. Allerdings wirken Defizite in diesem Bereich eher indirekt etwa über die Lesemenge (Artelt, Naumann & Schneider, 2010), während direkte Einflüsse auf das Leseverständnis als eher gering eingestuft werden können (McElvany et al., 2008). Als überholt gilt die These, dass leseschwache Kinder aufgrund allgemeiner Wahrnehmungsdefizite (z.B. im visuomotorischen Bereich) erkannt werden können (Herné & Löffler, 2014; Richter & Christmann, 2002). So tragen basale Wahrnehmungsprozesse, die unterhalb einer linguistischen Beschreibungsebene ansetzen, insgesamt nur wenig zur Erklärung von Unterschieden zwischen guten und schlechten Leserinnen und Lesern bei (Richter & Christmann, 2002). Im Folgenden werden die vier oben genannten Unterscheidungsmerkmale, die mit Defiziten im Leseverständnis assoziiert sind, beschrieben.

1) Defizite im Hörverstehen

Gemäß dem *Simple View of Reading*-Ansatz (Gough & Tunmer, 1986) wirken sich neben Defiziten bei Dekodierprozessen auch Hörverständnisprobleme auf das Leseverstehen aus (Marx & Jungmann, 2000, 82). Die meisten leseschwachen Kinder weisen allerdings in beiden Komponenten (Dekodierfähigkeiten *und* Hörverständnis) Defizite auf (Gough & Tunmer, 1986). Der genauere Blick auf die Entwicklung von mündlichem Sprachverständnis und Leseverständnis verdeutlicht, dass zunächst – besonders bei leseschwachen Kindern – die Fähigkeit zum mündlichen Sprachverständnis größer ist als jene im Leseverständnis, wobei sich das Verhältnis im Verlaufe der Zeit umkehrt (Klicpera et al., 2017). Leseschwache Kinder scheinen diese Entwicklung aber ohne spezielle Unterstützung nicht oder nur begrenzt mitzumachen (ebd.). So zeigt die empirische Studie von Rost und Hartmann (1992), dass Hör- und Leseverstehensleistungen bei der lesestärkeren Gruppe in der vierten Klasse stärker korrelieren als bei der schwächeren Gruppe; die leseschwachen Kinder bleiben in ihren Hör- wie auch Leseverstehensleistungen insgesamt deutlich hinter den lesestärkeren Kindern zurück.

2) Wortschatz

„Der Wortschatz ist einer der wichtigsten Prädiktoren für Erfolg beim Lesen“ (McElvany & Schneider, 2009, S. 160). Empirische Studien belegen, dass das Leseverhalten im Sinne der Lesehäufigkeit mit dem Umfang des Wortschatzes einhergeht und sich damit wiederum auf individuelle Unterschiede in den Leseleistungen auswirkt (McElvany et al., 2009). Allerdings scheint sich der Wortschatzumfang gemäß der Längsschnitt-Studie von Juska-Bacher et al. (2016) erst etwa ab der dritten Klasse auf der Ebene des Leseverstehens bemerkbar zu machen und weniger zu Beginn (erste/zweite Klasse) der Leseentwicklung, da Dekodierleistungen auf Wortebene zu diesem Zeitpunkt noch im Vordergrund stehen.

Im Verlaufe der Grundschule erweist sich der Umfang des Wortschatzes allerdings als immer zentraleres Unterscheidungskriterium hinsichtlich guten und schwachen Leseleistungen (Scheerer-Neumann, 2018; Schröder-Lenzen, 2013). In der Studie von Gold (2009) mit lese-schwachen Hauptschülern der sechsten Klasse erwies sich der Wortschatz als zweitwichtigster Prädiktor für das Textverstehen – hinter der Leseflüssigkeit als insgesamt dominierende Einflussgröße. Über die ersten drei Schuljahre hinweg lässt sich feststellen, dass die Bedeutung des Wortschatzes für das Leseverstehen über die Zeit stark zunimmt, und dass in der dritten Klasse mehr Varianz durch den Faktor Wortschatz erklärt wird als durch basale Lesekompetenzen (nonword reading) (Seigneuric & Ehrlich, 2005). Auch McElvany et al. (2009) belegen in ihrer Studie den deutlichen Zusammenhang zwischen Wortschatz und Leseverständnis in der vierten und sechsten Klasse. Allerdings wurde in dieser Untersuchung – wie auch in der Studie von Seigneuric und Ehrlich (2005) – der Fokus nicht speziell auf leseschwache Kinder gelegt, weshalb diese Beiträge lediglich als indirekte Belege für Schwächen in diesem Bereich dienen.

3) Syntaktisch-semantische Integration und Inferenzbildung

Mit Blick auf Studien aus dem englischen Sprachraum können Unterschiede zwischen unauffälligen und leseschwachen Kindern durch Probleme in der Inferenzbildung auf Satz- und Textebene und außerdem durch geringere Kompetenzen in der syntaktischen und semantischen Integration erklärt werden (Scheerer-Neumann, 2018). Auch gemäß der Forschungsübersicht von Richter, Isberner, Naumann und Kutzner (2012) bestehen keine Zweifel daran, dass individuelle Unterschiede in der Effizienz syntaktischer und semantischer Integrationsprozesse eng mit dem allgemein erreichten Leseverständnis zusammenhängen, und dass der Verarbeitungsaufwand mit wachsender syntaktischer und semantischer Komplexität von Sätzen, insbesondere bei schwächeren Leserinnen und Lesern, steigt. Da die Wortfolge miteinander in Beziehung gesetzt und in einer Gesamtstruktur integriert werden muss, sind Analysen der semantischen

und syntaktischen Relationen einzelner Satzelemente zentral (Lenhard et al., 2006). So zeigen sich Unterschiede zwischen guten und schlechten Leserinnen und Lesern etwa darin, dass sie irreführende Sätze („Sackgassen“-Sätze), die syntaktisch falsch strukturiert sind und in einem zweiten Schritt korrigiert werden müssen, unterschiedlich gut bewältigen (Lenhard et al., 2006, S. 15; Richter & Christmann, 2002, S. 40). Syntaktische und semantische Lesefehler werden von den Kinder allerdings oft gar nicht bemerkt und entsprechend auch nicht korrigiert (Schründer-Lenzen, 2013). Auch Kohäsionsmittel zur Verbindung zwischen Sätzen und Textstellen erweisen sich für jüngere und schwächere Leserinnen und Leser als Hürde (Lenhard et al., 2006). Damit sind entscheidende Voraussetzungen für das Verstehen komplexerer Texte nicht gegeben. Vermutet wird zudem, dass Probleme bei inferenziellen Prozessen auch mit einem geringeren Arbeitsgedächtnis zusammenhängen (Richter & Christmann, 2002).

Für das schlussfolgernde Denken bzw. Textverstehen ist das Bilden sogenannter lokaler und globaler Kohärenzen entscheiden, was wiederum mit dem Weltwissen und dem themenspezifischen Vorwissen assoziiert ist (Richter & Christmann, 2002; Scheerer-Neumann, 2018). Hier zeigen sich Defizite leseschwacher Kinder darin, dass sie sich wenig aktiv um ein Verständnis von Texten bemühen. Sie bleiben beim wörtlichen Verständnis stehen und ziehen kaum Schlussfolgerungen aus dem Gelesenen bzw. bilden keine Inferenzen (Klicpera et al., 2017; Richter & Christmann, 2002). Diese Beobachtung deckt sich mit der Untersuchung von Cain und Oakhill (1999), die zeigt, dass die leseschwachen bzw. jüngeren Kinder zwar von erneuter Darbietung des Textes profitieren und bei Nachfrage die entsprechenden Fragen beantworten konnten. Allerdings scheinen sie auf solche Hilfestellungen angewiesen, während gut lesende Kinder Strategiewissen automatisch anwenden und *von sich aus* Inferenzen bilden. Bei Kindern mit stark ausgeprägter Leseschwäche bleiben die Schwierigkeiten auch dann bestehen, wenn sie den Text vor sich haben und ihn zur Beantwortung der Fragen mit heranziehen können (Klicpera et al., 2017). Besonders ausgeprägt scheint das Problem, wenn sich Fragen auf Informationen beziehen, die nicht wörtlich im Text wiedergegeben sind sondern indirekt erschlossen werden müssen (ebd.).

4) Lesestrategien

Im weiterführenden Leseunterricht werden Lesestrategien für das Textverständnis immer wichtiger, und es zeigt sich zunehmend, dass gewisse Kinder zwar in der Lage sind, Texte für sich durchzulesen oder diese mehr oder weniger problemlos vorlesen, aber trotzdem größte Probleme haben, zentrale Argumentationslinien oder Hauptgedanken zu erschließen und die Absicht des Geschriebenen zu erfassen (Nix, 2010). Das angesprochene Problem einer fehlenden oder

ineffektiven Anwendung von Lesestrategien erweist sich demnach als weiteres Merkmal lese-schwacher Kinder (Gold et al., 2009; Hartmann, 2006; Kruse, 2006; Philipp, 2012a; Scheerer-Neumann, 2018). Es mangelt ihnen an Fähigkeiten, ihr Leseverstehen zu überwachen und in Situationen adäquate Strategien zur Erhöhung des Leseverstehens auszuwählen (Philipp, 2012a).

Der Forschungsüberblick von Müller (2015) beschreibt zusammenfassend, dass gute Leserinnen und Leser verschiedene (meta-)kognitive Lesestrategien wie das Zusammenfassen, Aktivieren von Vorwissen und Generieren von Fragen flexibel zur Steuerung und Regulation des Leseprozesses anwenden, während schwache Leserinnen und Leser diese nur selten spontan zur lokalen und globalen Kohärenzbildung nutzen. Gemäß den Ergebnissen der Längsschnitt-Studie von Vauras, Kinnunen und Kuusela (1994) zur Entwicklung der Lesestrategienutzung über die Zeit (dritte bis fünfte Klasse), ist von einem relativ stabilen Muster auszugehen: Über den gesamten Messzeitraum hinweg wies die Gruppe der leistungsschwachen Kinder hinsichtlich sämtlicher Strategien nur stagnierende Leistungen auf niedrigem Niveau auf.

Inwiefern sich Strategiewissen auf das Leseverständnis niederschlägt, berichten Müller (2015) und Philipp (2017) in ihren Forschungsüberblicken: Insgesamt konnte im Grundschulalter für zahlreiche Lesestrategien (z.B. Generieren von Fragen, Zusammenfassen von Textinhalten, Aktivieren von Vorwissen) ein positiver Zusammenhang mit dem Leseverständnis nachgewiesen werden. Der Zusammenhang zwischen Lesekompetenz und Strategiewissen zeigt sich dabei sowohl bei älteren als auch bei jüngeren Kindern. Das Strategiewissen wird aber aufgrund komplexerer Texte immer wichtiger. Gemäß PISA-Studie (Artelt et al., 2010, S. 104) liegen für Jugendliche (Ende neuntes Schuljahr) recht hohen Korrelationen (OECD-Mittelwert, $r = .50$) vor. In der finnischen Längsschnittstudie von Annevirta, Laakkonen, Kinnunen und Vauras (2007) wird dieser Zusammenhang auch bei jüngeren Kindern (Kindergarten bis dritte Klasse) belegt. Weitere Hinweise auf den Zusammenhang lassen sich auch in Interventionsstudien finden (siehe Übersicht Souvignier, 2009). In verschiedenen Lesestrategieprogrammen für die Grundschule lassen sich Effekte auf das Leseverständnis nachweisen (Bertschi-Kaufmann et al., 2007; Gold et al., 2009; Müller, 2015). Dabei wird die Wirksamkeit von Lesestrategietrainings für ältere Kinder bzw. für Kinder, die über gute basale Leseprozesse verfügen, als gewinnbringender eingeschätzt, da basale Lesekompetenzen eine Voraussetzung für die Anwendung sind (Gold et al., 2009; Müller, 2015; Rosebrock & Nix, 2012). Allerdings können auch bereits Kinder in der Grundschule Lese- und Verstehensstrategien anwenden (McElvany & Schneider, 2009).

2.9 Diagnostik bei Leseschwierigkeiten

In diesem Kapitel werden zunächst Möglichkeiten der Erfassung von Leseschwierigkeiten allgemein beschrieben. Anschließend werden in Kap. 2.9.1 diagnostische Verfahren, die im Rahmen der Prävention von Leseschwierigkeiten eingesetzt werden, erläutert. Ein Blick soll hier auch auf das Problem einer zuverlässigen Früherfassung von Risiken für Leseschwierigkeiten geworfen werden. Im Anschluss daran werden in Kap. 2.9.2 und 2.9.3 diagnostische Verfahren für die Teilbereiche der Lesekompetenz (basale Lesekompetenz, Leseflüssigkeit, Leseverständnis) beschrieben.

Möglichkeiten und Zeitpunkt der Erfassung von Leseschwierigkeiten

Wie entscheidend diagnostische Kompetenzen im Bereich der Leseförderung sein können, belegt u.a. die Studie von Behrmann & Souvignier (2013). Die Autoren konnten nachweisen, dass sich das Leseverständnis von Kindern signifikant verbessert, wenn Lehrpersonen sowohl über gute diagnostische Kompetenzen verfügen wie auch den Kindern viele Leistungsrückmeldungen geben. Insbesondere scheint eine prozessorientierte Diagnostik günstige Effekte auf die Leistungszuwächse von leseschwachen Kindern zu haben (Souvignier & Förster, 2011). Gemäß Günther (2007) ist die Prognose bei leseschwachen Kindern bei früher Identifikation und geeigneter Intervention sehr viel günstiger als wenn die Leseschwierigkeiten erst spät (dritte/vierte Klasse) erkannt werden.

Die Diagnose von Leseschwierigkeiten erfolgt gemäß Scheerer-Neumann (2018) am besten in verschiedenen Schritten: An erster Stelle steht die quantitative Ermittlung (Statusdiagnostik) der entsprechenden Teilleistung mithilfe standardisierter und normierter Tests. In einem nächsten Schritt können entwicklungsbezogene Analysen durchgeführt werden (Verlaufsdiagnostik). Somit lassen sich grundsätzlich zwei verschiedene Arten der Diagnostik unterscheiden, nämlich die „Status-“ und die „Verlaufsdiagnostik“ (Lenhard & Lenhard, 2017, S. 174). Die *Statusdiagnostik* (und teilweise auch Verlaufsdiagnostik) verläuft in aller Regel normorientiert und erfolgt mittels standardisierter Testverfahren, die einzelne Bereiche der Lesekompetenz wie basales Lesen, Leseflüssigkeit, Leseverständnis erfassen (Scheerer-Neumann, 2018). Individuell erreichte Leistungen werden meist in Form von Prozentrang (PR) und T-Wert angegeben (Herné & Löffler, 2014). Mit dem PR lässt sich ablesen, wo ein Kind hinsichtlich seiner Leistung vor dem Hintergrund der Normstichprobe steht, indem mit dem Anteil an Kindern, die gleich gute oder schlechtere Leistungen erzielen, verglichen wird. „Von unterdurchschnittlichen Leistungen spricht man im Allgemeinen, wenn sie unter einem PR von 15 liegen, wobei hier auch manchmal auch ein PR von 25 angesetzt wird, so dass der Bereich zwischen PR 15

und PR 25 als Grenzbereich angesehen werden kann“ (Marx, 2007b, S. 81). Das Ziel *verlaufsdiagnostischer Verfahren* besteht darin, Kindern in häufigen und regelmäßigen Abständen formative Rückmeldungen zu ihrem Leistungsstand zu geben. Dafür muss eine große Anzahl unterschiedlicher Aufgaben gelöst werden (Lenhard & Lenhard, 2017).

Die diagnostischen Möglichkeiten zur Erfassung der Lesekompetenz – zumindest ab dem Schulalter – haben sich innerhalb des letzten Jahrzehnts deutlich verbessert (Schneider, 2017). Es liegen inzwischen zahlreiche normierte diagnostische Verfahren vor, mit denen sich Leseschwierigkeiten in verschiedenen Teilbereichen zuverlässig erkennen lassen (siehe Übersichten: Lenhard & Lenhard, 2017; Marx, 2007b; Mayer, 2016; Schneider, 2009/2017). Bereits ab Ende der ersten Klasse können Aussagen zum Vorliegen möglicher Risiken für spätere Leseschwierigkeiten gemacht werden (Diehl & Hartke, 2012). Wie in den vorangegangenen Kapiteln beschrieben, weisen leseschwache Kinder allerdings bereits wenige Wochen nach Einschulung einen deutlichen Rückstand auf gleichaltrige Kinder auf. Die erste Messung einer Teilstichprobe der Wiener Längsschnittstudie erfolgte neun Wochen nach Schulbeginn (Klicpera & Gasteiger-Klicpera, 1993), wobei allerdings eine Orientierung an Normen vor Ende der ersten Klasse kaum möglich ist, da der bis zur Testdurchführung vorgenommene Erstleseunterricht Auswirkungen auf die Leseentwicklung hat (Marx, 2007b). Vorläuferfertigkeiten können allerdings bereits im Kindergarten mit entsprechenden normierten Tests erfasst, womit sich zumindest *Risiken* für spätere Leseschwierigkeiten beurteilen lassen (Mayer, 2016).

Ergänzend zur normorientierten Status- und Verlaufsdagnostik bestehen systematische *Beobachtungsaufgaben im Schulalltag*, die sich weniger an Normen als vielmehr an der Qualität der Leseleistungen orientieren und zudem die individuelle Förderung stärker in den Blick nehmen (Herné & Löffler, 2014). Für die kontinuierliche Beobachtung der Leseentwicklung und des Leseverhaltens eignen sich nicht-normierte Verfahren wie die Anwendung von Beobachtungsrastern (Bertschi-Kaufmann, 2011). Solche Beobachtungen im Schulalltag beziehen sich auf sichtbare Merkmale des Leseverhaltens (ebd.). Inzwischen liegen für den Schulalltag zahlreiche systematische diagnostische Verfahren, die zudem theoretisch abgestützt sind, vor (u.a. "Lesen. Das Training": Bertschi-Kaufmann et al., 2007; "ILeA": Liebers & Prengel, 2010; "Lesestufen": Niedermann & Sassenroth, 2017). Die Ergebnisse solcher Verfahren können zwar nicht zur Diagnose einer Leseschwäche führen. Sie können aber den Blick für die Notwendigkeit weiterführender Abklärungen öffnen und zudem aufzeigen, an welchen Aufgaben Kinder scheitern und wo die Förderung entsprechend ansetzen muss.

2.9.1 Diagnostik der phonologischen Informationsverarbeitung

Bei der Prävention von Leseschwierigkeiten spielt die Diagnose von Schwierigkeiten bei den Vorläuferfertigkeiten (phonologische Bewusstheit / phonologisches Arbeitsgedächtnis / Benennungsgeschwindigkeit) eine bedeutsame Rolle. Zu allen drei Komponenten der phonologischen Informationsverarbeitung liegen diagnostische Verfahren vor, mit denen Risiken für spätere Leseschwierigkeiten bereits im Kindergartenalter erfasst werden können (siehe Überblicke: Marx, 2007b; Mayer, 2016). Die frühe Erkennung von Kindern mit Risiken ist allerdings mit dem Problem verbunden, dass sich eine Over- oder Underprediction nicht ganz vermeiden lässt. „Overprediction“ (falsch positiv) meint die Klassifizierung eines Kindes als Risikokind, das sich später unauffällig entwickelt, während „Underprediction“ (falsch negativ) genau das Umgekehrte meint (Mayer, 2016, S. 122f.). Letzteres ist im Hinblick auf eine verpasste Chance auf frühe Förderung sicherlich das größere Problem.

Ein frühzeitiges Erkennen von Risikokindern ist unabdingbar, da bereits ab der dritten Klasse selbst gut konzipierte Interventionen nur noch unspektakuläre Effekte zeigen. Der beachtliche Leistungsrückstand zu nicht leseschwachen Kindern kann bei später Diagnose oft nicht mehr wesentlich aufgeholt werden (Hartmann, 2007). Eine zuverlässige Diagnose im Vorschulalter ist allerdings schwierig, da der Unterricht (Aufbau von Phonem-Graphem-Korrespondenz) die frühen Kompetenzen des Leseerwerbs stark beeinflusst und in diesem Sinne „ausgleichend“ wirken kann, was sich besonders bei der phonologischen Bewusstheit bemerkbar macht (Klicpera et al., 2017; Marx, 2007b). Eine zuverlässige Vorhersage lässt sich in der Regel eher erreichen, wenn mehrere relevante Vorläuferfertigkeiten erfasst werden. Als bedeutsam gelten das RAN und die phonologische Bewusstheit (Mayer, 2016). Zur prädiktiven Validität verschiedener Verfahren liegen u.a. Analysen von Marx (2007b) und Mayer (2016) vor, die insgesamt ein eher kritisches Bild auf die Vorhersagbarkeit von Lese-Rechtschreibschwächen im Vorschulalter werfen. Verschiedene Verfahren erfassen zwar relevante Vorläuferfertigkeiten, die über dem Zufall liegende Vorhersagen von Lese-Rechtschreibschwächen erlauben (Marx, 2007b). Allerdings ist die Treffsicherheit der Früherkennung unbefriedigend, weshalb die Verfahren bezüglich prognostischer Validität nur bedingt geeignet sind (Mayer, 2016). Verschiedentlich wird deshalb vor zu hohen Erwartungen an eine vorschulische oder frühe schulische Identifikation von Risikokindern gewarnt, da vielfältige Einflüsse und individuelle Einflussfaktoren auf den Schriftspracherwerb wirken und einer treffsicheren vorschulischen Prognose entgegenstehen (Marx, 2007b; Mayer, 2016; Valtin, 2010). Am präzisesten sind Vorhersagen für Kinder möglich, die im Sinne eines „double-deficit“ sowohl bei der Benennungsgeschwindigkeit als auch in der phonologischen Bewusstheit unterdurchschnittliche Werte erreichen

(Mayer, 2014, S. 396). Im Allgemeinen scheint die Identifizierung von Nicht-Risiko Kindern, die später mit hoher Wahrscheinlichkeit *keine* Leseschwäche entwickeln, besser möglich als das Erkennen von Risiko-Kindern (Martschinke, Kirschhock & Frank, 2018; Mayer, 2016).

Erfassung der phonologischen Informationsverarbeitung

Die bekanntesten standardisierten Verfahren, die jeweils verschiedene Voraussetzungen des Schriftspracherwerbs prüfen und teilweise bereits vor Beginn der ersten Klasse erstmalig eingesetzt werden können, sind das „Bielefelder Screening“ [BISC] (Jansen, Mannhaupt, Marx & Skowronek, 2002), das „Münsteraner Screening“ [MÜSC] (Mannhaupt, 2006), der „Test zur Erfassung der phonologischen Bewusstheit und der Benennungsgeschwindigkeit“ [TEPHOBE] (Mayer, 2011), der Test „Basiskompetenzen für Lese-Rechtschreibleistungen“ [BAKO 1-4] (Stock, Marx & Schneider, 2017) sowie der „Rundgang durch Hörhausen“ (Martschinke et al., 2018). Die genannten Testinstrumente bzw. Screenings erfassen verschiedene Aspekte der phonologischen Informationsverarbeitung. Sie sind bereits im Kindergarten einsetzbar und erfassen teilweise zusätzliche relevante Aspekte wie die Benennungsgeschwindigkeit und das phonologische Arbeitsgedächtnis sowie die Buchstabenkenntnis.

Phonologische Bewusstheit: Sie wird typischerweise über das Identifizieren, Segmentieren, Synthetisieren und Manipulieren von Silben, Onset-Rime und Phonemen erfasst (Kap. 2.5.2). Typische Aufgaben können exemplarisch folgendermaßen aussehen:

- 1) Erkennen von Reimen (Beispiel aus BISC: Kind – Wind oder Kind – Stuhl).
- 2) Kategorisieren von An- und Auslauten (Beispiel aus BAKO 1-4: Satz – Sohn – Ton – sehr).
- 3) Phonemanalyse (Beispiel aus Rundgang durch Hörhausen: Welches Wort ergibt sich, wenn man /a:/-/s/-/t/ verbindet).
- 4) Lautsynthese (Beispiel aus TEPHOBE: Welches Bild passt zu /t/ /i/ /f/ → Bilder von: Tisch / Fisch, Tasche, Schiff).
- 5) Wortumkehraufgabe (Beispiel aus Rundgang durch Hörhausen: Welches Wort ergibt rückwärts <ENAP>).
- 6) Vokalersetzung (Beispiel aus BAKO 1-4: Jedes im Testwort vorkommende /a:/ soll durch ein /i:/ ersetzt werden: <Kaba> → <Kibi>).

Benennungsgeschwindigkeit: Die Benennungsgeschwindigkeit wird üblicherweise durch das schnelle Abrufen von Symbolen aus dem Langzeitgedächtnis erfasst. Im BISC (Jansen et al., 2002) und MÜSC (Mannhaupt, 2006) erfolgt dies durch das schnelle Benennen von Farben verschiedener farbiger Objekte sowie das Benennen der Farben bei vorliegenden schwarz/weiß Objekten

(Tomate, Salat, Banane und Pflaume). Im TEPHOBE erfolgt die Erfassung über das schnelle Benennen von Farben, Zahlen oder Objekten (Mayer, 2016).

Phonologisches Arbeitsgedächtnis: Die Effizienz des phonologischen Arbeitsgedächtnisses wird typischerweise bereits im Vorschulalter erfasst, indem Kindern vier- bis fünfsilbige Pseudowörter (z.B. <risolamu>) vorgesprochen werden, die sie anschließend nachsprechen müssen (Jansen et al., 2002; Mayer, 2016). Im MÜSC wird eine Wortfolge von vier Wörtern vorgesprochen, die im Anschluss daran in verschiedenen vorgegebenen Bildfolgen wiedererkannt werden muss (Mannhaupt, 2006).

Buchstabenkenntnis: Neben Aspekten der phonologischen Informationsverarbeitung können bereits im Vorschulalter oder bei Schuleintritt erste Buchstabenkenntnisse geprüft werden. Dies erfolgt z.B. über das schnelle Benennen von Buchstaben, die in Buchstabenlisten in einer zufälligen Reihenfolge vorliegen (Klicpera et al., 2017). Zu einem späteren Zeitpunkt, etwa Ende der ersten Klasse, wird die Buchstabenkenntnis etwa im IEL-1 dadurch erfasst, dass in vorgegebenen Wörtern mit und ohne Bildunterstützung die fehlenden Grapheme (lautgetreu) ergänzt werden müssen (z.B. A__pel, Au__o, M__nn) (Diehl & Hartke, 2012). Damit wird zusätzlich das korrekte phonologische Rekodieren erfasst (ebd.).

2.9.2 Diagnostik der basalen Lesekompetenzen und Leseflüssigkeit

Im Laufe der letzten Jahre wurden zahlreiche standardisierte und normierte Lesetests entwickelt oder überarbeitet. Sie wurden hinsichtlich psychometrischer Qualitäten überprüft, und bei der Konstruktion orientierten sich die Verfassenden an aktuellen Leseprozessmodellen (Klicpera et al., 2017). So liegt gemäß Einschätzung von Mayer (2016) im deutschsprachigen Raum „eine ausreichende Anzahl durchdachter Verfahren zur Früherkennung und Erfassung zentraler schriftsprachlicher Kompetenzen“ (S. 153) vor. Die Kritik an den Verfahren besteht allenfalls darin, dass die Normen teilweise regional eingeschränkt sind und sich deshalb nicht ohne weiteres auf andere Populationen übertragen lassen (Lenhard & Lenhard, 2017). In Anlehnung an die Übersichten zu aktuellen Verfahren zur Diagnose des Worterkennens bzw. zu Lesesicherheit und Lesegeschwindigkeit (Klicpera et al., 2017; Lenhard & Lenhard, 2017; Mayer, 2016; Scheerer-Neumann, 2018) sollen hier exemplarisch einige Formate der Erfassung aufgeführt werden. Es handelt sich im Folgenden um Tests, die klassischen psychometrischen Gütekriterien genügen und im Zeitraum erste bis Ende vierte Klasse (teilweise aber auch in höheren Klassen) eingesetzt werden können.

Diagnostik des basalen Wort-, Satz- und Textlesens: Lesegeschwindigkeit und Lesegenauigkeit

Grundlage der Tests zum Wortlesen sind häufig 1-Minuten-Lesetests, bei denen meist innerhalb eines festgelegten Zeitraums so viele Wörter wie möglich richtig zu lesen sind (Klicpera et al., 2017). Aktuelle normierte Tests, die nach diesem Prinzip u.a. die Dekodier- und/oder Dekodiergeschwindigkeit erfassen, sind die „Würzburger Leise Leseprobe“ [WLLP-R] (Schneider, Blanke, Faust & Küspert, 2011), das „Salzburger Lese-Screening“ [SLS 2-9] (Mayringer & Wimmer, 2014), der „Salzburger Lese- und Rechtschreibtest“ [SLRT-II] (Moll & Landerl, 2014), der „Zürcher Lesetest-II“ [ZLT-II] (Petermann & Daseking, 2019), der Test „Lernfortschrittdiagnostik Lesen“ [LDL] (Walter, 2010) sowie das „Inventar zur Erfassung der Lesekompetenz im 1. Schuljahr“ [IEL-1] (Diehl & Hartke, 2012). Außer dem SLS 2-9 können alle genannten Verfahren ab der ersten Klasse zu festgelegten Zeitpunkten (Normierung) durchgeführt werden. In den Tests werden basale Lesekompetenzen mit teilweise recht unterschiedlichen Formaten abgefragt, was u.a. damit zu tun hat, dass sie als Gruppen- oder Einzeltest vorliegen.

Lesegenauigkeit (Pseudowörter): Diese wird in einem klassischen 1-Minuten-Lesetest etwa im SLRT-II über das laute Vorlesen von Pseudowortlisten erfasst (Moll & Landerl, 2014). Gemessen wird, wie viele Pseudowörter innerhalb einer Minute korrekt gelesen werden. Beim Lesen von Pseudowörtern zeigen sich Syntheseprobleme besonders deutlich, da im semantischen Lexikon kein Eintrag vorliegt. Mit diesem Format wird ausschließlich das alphabetische, phonologische Rekodieren überprüft und damit auch die Genauigkeit, mit der Grapheme in Phoneme übersetzt werden (Klicpera et al., 2017; Scheerer-Neumann, 2018). Solche Items eignen sich insbesondere zur Erfassung von Kindern, bei denen der Verdacht besteht, dass sie ratend lesen. Ausgehend davon, dass ein Text in der Regel nur dann verstanden wird, wenn die Zahl der Lesefehler unter 10% liegt, ist das Textverstehen bei leseschwachen Kindern auch durch die Lesefehler beeinträchtigt (Tacke, 2011, S. 143). Deshalb ist neben der Lesegenauigkeit auch die Lesegeschwindigkeit zu berücksichtigen. Das Verringern der Lesefehler hat zudem positive Auswirkungen auf die Lesezeit (ebd.).

Wort-Lesegeschwindigkeit: Analog zum Vorgehen beim Lesen von Pseudowörtern können auch echte Wörter in Wortlisten vorgelegt werden, wie z.B. beim SLRT-II, dem IEL-1 oder dem ZLT-II. Bei diesem Vorgehen können Kinder auf einen Eintrag im orthographischen Lexikon oder ggf. auch auf einen bereits erworbenen Sichtwortschatz zurückgreifen (Scheerer-Neumann, 2018). Die Dekodiergeschwindigkeit wird abgefragt (z.B. WLLP-R), indem einem geschriebenen Wort jeweils vier Bildalternativen gegenübergestellt werden, von denen das

korrespondierende Bild anzustreichen ist (Schneider et al., 2011). Beim IEL-1 werden die Lesegeschwindigkeit und die Lesegenauigkeit durch Items erfasst, bei denen jeweils ein Bild mit jeweils einem Wort aus einer Auswahl von verschiedenen, sich ähnelnden Zielwörtern zugeordnet werden muss (z.B. Bild einer Rose → Nase, Hose, Hase, Rose) (Diehl & Hartke, 2012, S. 18). Damit ist das Wort-Bild-Verhältnis im Vergleich zur WLLP-R genau umgekehrt, und die Schwierigkeit wird durch die Wortähnlichkeit wohl etwas höher.

Leseflüssigkeit auf Satz- und Textebene: Die Leseflüssigkeit wird überprüft, indem Sätze oder ganze Texte vorliegen, die entweder laut vorgelesen (z.B. LDL) oder leise gelesen werden (SLS 2-9). Beim LDL werden zu verschiedenen Zeitpunkten Parallelformen von Texten mit ähnlicher Textschwierigkeit laut vorgelesen, wobei die Anzahl korrekter Wörter pro Minute bestimmt werden muss (Walter, 2013). Beim SLS 2-9 werden Sätze leise gelesen, und es muss entschieden werden, ob die jeweilige Aussage richtig oder falsch ist (Mayringer & Wimmer, 2014). Dabei werden an die inhaltliche Beurteilung nur geringe Anforderungen gestellt (z.B. „Bananen sind blau“ → richtig/falsch), da lediglich sichergestellt werden soll, dass auch wirklich *gelesen* wird (Scheerer-Neumann, 2018, S. 82).

Um Hinweise für mögliche weitere Abklärungen zu erhalten, sind auch informelle Verfahren, die im Schulalltag ohne großen Aufwand durchgeführt werden können, geeignet (Klicpera et al., 2017; Rosebrock et al., 2011). Beispielsweise lesen Kinder während einer Minute einen nicht zu schwierigen Text *prima Vista* vor. Während des Lesens werden die unkorrigierten und falsch gelesenen Wörter markiert (Rosebrock et al., 2011). Die Beurteilung der Lesegeschwindigkeit ist dabei klassenstufenabhängig und liegt z.B. für die dritte Klasse bei einer Rate von 45 bis 60 richtig gelesener Wörter pro Minute (Klicpera et al., 2017). Um von *flüssigem Lesen* zu sprechen, gilt ein Richtwert von ca. 100 korrekt gelesenen Wörtern pro Minute (Rosebrock et al., 2011). Mit den genannten Verfahren wird indirekt auch der Automatisierungsgrad auf Wortebene erfasst, da ein buchstabierendes Erlesen von Wörtern wesentlich mehr Zeit braucht und damit auch die Lesegeschwindigkeit beeinflusst (ebd.).

Ergänzend zu den genannten Tests kann für eine differenzierte qualitative Analyse von Lesefehlern (auf Wort-, Satz und Textebene wie auch für das Leseverständnis) die „Hamburger Leseprobe“ [HLP] (May & Arntzen, 2003) eingesetzt werden. Die HLP umfasst Leseraufgaben, die insbesondere für die Beobachtung von schwachen Leserinnen und Lesern geeignet sind (May & Arntzen, 2003, 3). Mit dem Verfahren können genauere qualitative Analysen über die *Art* von Lesefehlern gemacht werden, indem z.B. die Qualität des synthetisierenden Lesens, die Art der Verlesungen oder das Korrigierverhalten in einer dreistufigen Skala beurteilt werden.

Das Verfahren wurde allerdings für einige Teile nicht normiert, und es liegen keine Angaben zu psychometrischen Testgütekriterien vor.

Prosodie: Neben der Geschwindigkeit und der Genauigkeit werden im Rahmen der Diagnostik der Leseflüssigkeit auch prosodische Aspekte erfasst. Allerdings gibt es bei den formellen diagnostischen Verfahren aktuell noch keinen Test, der prosodische Kompetenzen objektiv ermittelt (Scheerer-Neumann, 2018). Die Intonation kann entsprechend nicht direkt über Lautleseprotokolle erfasst werden. Sie erfolgt meist zusätzlich im Rahmen von 1-Minuten-Lesetests, z.B. über eine subjektive Einschätzung mittels „Checkliste zur Erfassung der Intonation nach Pinnell (1995)“. Mit Hilfe dieser Rating-Skala, die u.a. für die Forschungsprojekte von Nix (2011) und Rosebrock et al. (2011) leicht modifiziert wurde, werden „die Anzahl der jeweils zusammen gelesenen Wörter, die syntaktische Angemessenheit dieser Wortgruppierungen sowie die expressive Interpretation beim Vorlesen“ (Scheerer-Neumann, 2018, S. 92) in einer vierstufigen Skala beurteilt. Anhand der Skala soll ermittelt werden, ob Kinder noch Wort für Wort lesen (Niveau 1), bereits in Zweier-Wortgruppen lesen (Niveau 2), teilweise in größeren Wortgruppen und angemessen prosodisch lesen (Niveau 3) oder weitgehend in größeren, semantisch sinnvollen Worteinheiten lesen (Niveau 4) (Rosebrock et al., 2011).

2.9.3 Diagnostik des Leseverständnisses

Auch für das Leseverständnis liegen verschiedene normierte und standardisierte Gruppentests vor (Klicpera et al., 2017). In aktuellen Übersichten (Klicpera et al., 2017; Lenhard & Lenhard, 2017; Mayer, 2016; Scheerer-Neumann, 2018) werden insbesondere die folgenden Tests besprochen: Der „Leseverständnistest für Erst- bis Sechstklässler“ [ELFE 1-6] (Lenhard et al., 2006), der „Leseverständnistest für Erst- bis Siebtklässler“ [ELFE II] (Lenhard, Lenhard & Schneider, 2020), der „Hamburger Lesetest für 3. und 4. Klassen“ [HAMLET 3-4] (Lehmann, Peek & Poerschke, 2006), die „Verlaufsdiagnostik sinnverfassendes Lesen“ [VSL] (Walter, 2013). Verschiedene weitere Tests enthalten neben Items zur Erfassung basaler Lesekompetenzen zusätzliche Items zum Leseverständnis, wie etwa der bereits erwähnte IEL-1 (Diehl & Hartke, 2012) oder der Test „Prozessbezogene Diagnostik von Lesefähigkeiten im Grundschulalter“ [ProDi-L] (Richter, Naumann, Isberner, Neeb & Knoepke, 2017).

Aufgrund des unterschiedlichen Theoriebezugs in den einzelnen Tests fallen auch die Formate, mit denen das äußerst heterogene Konstrukt des Leseverständnisses erfasst wird, verschieden aus. Während im IEL-1 und in der VSL ein deutlicher Bezug zum *Simple View of Reading*-Ansatz hergestellt wird, werden im ProDi-L, im ELFE 1-6 und im HAMLET 3-4 Theorien zur grammatikalischen Verarbeitung von Sätzen, Theorien zu mentalen Modellen (Integration von

Vorwissen) sowie Theorien zum inferenziellen Lesen ins Zentrum gerückt. Zur Erfassung des Leseverständnisses werden vor allem zwei Verfahren genutzt: die „Close Procedure“, bei der ein passendes Wort in eine Satzlücke eingefügt werden muss oder die „Maze Procedure“ (= Irrgarten, Labyrinth), bei der aus einer Alternative von Wörtern das richtige eingesetzt werden muss. Beim zweiten Verfahren müssen nach Lektüre eines Textes implizite oder explizite Fragen beantwortet werden (Scheerer-Neumann, 2018, S. 104; Walter, 2013, S. 12). Exemplarisch werden im Folgenden typische Formate zur Erfassung des Leseverständnisses auf Wort-, Satz- und Textebene besprochen.

Wortebene: Wie weiter oben unter der Rubrik „Wort-Lesegeschwindigkeit“ beschrieben, beinhalten alle genannten Tests auf Wortebene Worterkennungs-Items. Das heißt, dass aufgrund eines Bildes aus einer Wortauswahl der entsprechende Begriff gewählt werden muss. Bezüglich dieses Verfahrens wird unter mehr oder weniger explizitem theoretischen Bezug auf Zwei-Wege-Modelle des Wortlesens argumentiert, dass das Leseverständnis wesentlich vom Erkennen der Bedeutung einzelner Wörter abhängt (Lenhard et al., 2006; Richter et al., 2012; Walter, 2013).

Ganz anders wird der Zugriff auf die Wortbedeutung im ProDi-L (Richter et al., 2012) erfasst, bei dem semantische Kategorien gebildet werden müssen. Das heißt, dass ein Kategoriename (z.B. *Obst*) auditorisch präsentiert wird und einem Testwort (z.B. *Apfel*), das unmittelbar darauf visuell auf dem Bildschirm erscheint, zugeordnet werden muss. „Die Aufgabe der Testperson besteht darin, zu entscheiden, ob das visuell präsentierte Wort einen Unterbegriff des Konzepts darstellt, das durch den Kategoriennamen angesprochen wird“ (ebd., S. 316).

Satzebene: Auf dieser Ebene spielen syntaktische und semantische Teilprozesse eine wichtige Rolle, und sogenannte „Sackgassen-Effekte“ können im Rahmen von maze-Verfahren genutzt werden (Lenhard et al., 2006, S. 15). Ebenso wie im weiter oben gezeigten Beispiel werden im ELFE-1-6 Sätze vorgelegt, bei denen aus einer Wortauswahl der richtige Begriff bestimmt werden muss (hier Konjunktionen, z.B. *Lea spielt, [anstatt / nachdem / dass / bevor / damit] zu lernen*) (ebd., S. 15). Nach ähnlichem Prinzip funktionieren auch die Items der ProDi-L (Richter et al., 2012, S. 317), bei der Sätze hinsichtlich ihrer „syntaktischen Wohlgeformtheit“ beurteilt werden müssen (z.B. *Die Autos sind über die Straße fahren*). Zur Erfassung der semantischen Integration werden im selben Test Satzverifikationsaufgaben mit wahren oder falschen begrifflichen Aussagen vorgelegt, die eine einfache syntaktische Struktur aufweisen und vertraute Wörter enthalten (z.B. *Schnecken sind schnell*). „Diese Aufgabe erfordert die Extraktion eines kohärenten Satzsinns und (...) eine Überprüfung des Wahrheitsgehalts anhand des eigenen Weltwissens“ (ebd., S. 317).

In einem etwas anderen Format wird im IEL-1 im Prinzip ebenfalls diese Art von Satzverständnis gemessen. Hier werden drei Aufgabentypen verwendet, bei denen Kinder bildlich dargestellte Situationen und eine Auswahl zugehöriger Sätze auf ihre Richtigkeit hin beurteilen müssen. Damit wird erfasst, inwieweit Kinder „Einzelnformationen auf der Basis von Wörtern in größeren Sinneinheiten erfassen können“ (Diehl & Hartke, 2012, S. 18).

Textebene: Zur Überprüfung des Textverständnisses werden im HAMLET 3-4 (Lehmann et al., 2006) und im ELFE 1-6 (Lenhard et al., 2006) Texte verschiedener Komplexität vorgelegt, zu denen im Anschluss ans Lesen Fragen beantwortet werden. Dabei werden unterschiedliche Niveaustufen des Textverständnisses erfasst. Auf einfachster Ebene handelt es sich um das Auffinden isolierter Informationen, die im Text explizit genannt werden. Die Herstellung anaphorischer Bezüge (Rückverweise auf bereits genannte Satzteile) durch Kohäsionsmittel erfordert bereits komplexere kognitive Fähigkeiten (Mayer, 2016). Auf komplexestem Niveau wird schließlich das inferenzielle Lesen erfasst, d.h. das schlussfolgernde Lesen durch nicht explizit genannte Informationen (ebd.).

Auch im IEL-1 wird das inferenzielle Lesen geprüft, einmal mit demselben Format wie oben erwähnt, einmal mit Bildunterstützung. Dabei muss ein kurzer Text mit einem Bild aus einer Bild-Auswahl verbunden werden, wobei der Text inhaltlich nur zu *einem* Bild passt, während die anderen auch nicht übereinstimmende Informationen enthalten (Diehl & Hartke, 2012).

Ganz anders werden in der VSL und in der ProDi-L Aspekte des sinnverstehenden Textlesens erfasst. In der VSL wird das sinnverstehende Lesen unter Verwendung der Maze-Prozedur erhoben, indem Kinder während vier Minuten einen Text lesen und bei etwa jedem siebten Wort den korrekten Begriff aus drei Auswahlwörtern bestimmen müssen (Walter, 2013, S. 16) (z.B. *Danach [erschüttern, füttern, essen] sie die Hühner und die dicken [Schweine, Beine, Schwerter]*). Erfasst werden damit neben dem Dekodierprozess auch syntaktisches und semantisches Wissen sowie Kompetenzen auf Ebene Wortschatz und Weltwissen (ebd.). Schließlich basieren die Aufgaben aus der ProDi-L auf der Effizienz lokaler Kohärenzbildungsprozesse und der Fähigkeit allgemeines Weltwissen zu integrieren (Richter et al., 2017). Kinder müssen aufgrund kurzer „Geschichten“ entscheiden, ob Sätze zueinander passen bzw. ob der zweite Satz die Geschichte richtig weitererzählt (z.B. *Katrin muss ins Krankenhaus. Sie ist nämlich ganz gesund* → richtig/falsch) (ebd., S. 20).

2.10 Zusammenfassung und Fazit zu relevanten Wissensgrundlagen im Bereich Lesen und Leseschwierigkeiten

Die vorangehenden Kapitel zeigen die Komplexität des Wissensbereichs „Lesen“ und machen deutlich, dass Lehrpersonen auf verschiedenen Ebenen kompetent sein müssen, um Leseprozesse und diesbezügliche Beeinträchtigungen adäquat beurteilen zu können. Zum einen müssen Komponenten der Lesekompetenz wie basale Lesefertigkeiten, Leseflüssigkeit und Leseverständnis analysiert und zum anderen ihr Zusammenwirken verstanden werden. Eine grundlegende Erkenntnis ist, dass gelingende hierarchieniedrige Prozesse Voraussetzung für die Entwicklung hierarchiehoher Prozesse sind.

Weiter müssen für einen gelingenden Leseanfangsunterricht Einsichten in Entwicklungs- und Stufenmodelle erlangt werden. Entwicklungsmodelle liefern Erklärungen dazu, welche Prozesse im Moment des Lesens ablaufen (u.a. Zwei-Wege-Modell), Stufenmodelle beschreiben die längerfristige Entwicklung vom ersten vorschulischen Kontakt mit Schriftsprache bis zum kompetenten Lesen am Ende der Schulzeit. Auch das Wissen über relevante Vorläuferfertigkeiten (z.B. phonologische Informationsverarbeitung), die den Leseerwerbsprozess beeinflussen, gehört zum Standardwissen einer Lehrperson. Das Wissen um ineinander übergehende dominante Phasen (logographische, alphabetische und orthographische Phase) ist entscheidend, um den ungestörten und vor allem auch den erschwerten Leseerwerb erkennen zu können.

Weiter sind für das Verständnis dieser Prozesse und Entwicklungen linguistische Grundlagenkenntnisse zur jeweiligen Schriftsprache notwendig. Durch sprachstrukturelle Gegebenheiten verursachte Leseschwierigkeiten können dadurch besser verstanden werden. Damit gehen auch Einsichten dazu einher, welche textseitigen Merkmale das Lesen von Wörtern, Sätzen und Texten erschweren oder erleichtern.

Eine weitere wesentliche Aufgabe von Lehrpersonen besteht darin, Leseschwierigkeiten zu erkennen und Diagnosen zu stellen. Kenntnisse über spezifische Symptome von Leseschwierigkeiten sind dafür unabdingbar. Ergänzend dazu müssen die diagnostischen Möglichkeiten bekannt sein, und sie müssen für den gezielten Informationsgewinn bei Schwierigkeiten in den Bereichen phonologische Informationsverarbeitung, basale Lesekompetenzen, Leseflüssigkeit und Leseverständnis eingesetzt werden können.

3 Professionelle Kompetenz und Professionswissen von Lehrpersonen

3.1 Theoretische Bestimmung von professioneller Kompetenz und Professionswissen bei Lehrpersonen

Kompetenzorientierung in pädagogischen Berufen ist nicht erst seit heute ein Thema. Dies zeigen etwas ältere Publikationen wie etwa die Ordnungssysteme von Shulman (1986; 1987) und Bromme (1997), auf die sich nahezu alle Arbeitsgruppen berufen (Krauss et al., 2017b). Überblickt man die Forschung und die Publikationen der letzten Jahre, so zeigt sich ein großes Bestreben, Kompetenzmodelle auszudifferenzieren und auch *empirisch* zu begründen (u.a. Blömeke et al., 2010b; Kunter et al., 2011a; Lindmeier, 2011). Kern dieser Anliegen ist die Operationalisierung professioneller Kompetenz von Lehrpersonen. Ein Überblick zu verschiedenen Facetten und wiederkehrenden Begriffen soll Klarheit schaffen, wovon im Allgemeinen gesprochen wird, wenn das Konstrukt „Kompetenz“ und assoziierte Begriffe (Professionswissen, Professionalität, Expertise) verwendet werden. Im Folgenden werden Aspekte der professionellen Kompetenz dargestellt sowie Typologien bzw. Modelle skizziert, die in richtungsweisenden Studien verwendet wurden. Zentrales Anliegen ist es, das vorliegende Projekt in der aktuellen theoretischen Diskussion zu situieren und wesentliche Merkmale des Kompetenzbegriffs und des Begriffs „Professionswissen“, die für die Entwicklung des Befragungsinstruments verwendet wurden, zu umschreiben.

3.1.1 Kompetenz und Professionswissen

Verortung von Professionswissen im Konstrukt der Kompetenz

Die Zunahme an Forschung zu Lehrer- und Lehrerinnenkompetenzen ist u.a. durch die Erkenntnis bedingt, dass die Qualität von Unterricht maßgeblich von der täglich geleisteten Arbeit der Lehrpersonen abhängt (König, 2015). Somit gehört die Kompetenz von Lehrpersonen zu den elementaren Bedingungsfaktoren des Lernens von Schülerinnen und Schülern (Maag Merki & Werner, 2014). Bezüglich des Kompetenzbegriffs bezieht sich die empirische Forschung wiederkehrend auf den Begriff von Weinert (2001), bei dem Kompetenzen von Lehrpersonen als kognitive Fähigkeiten und Fertigkeiten verstanden werden, um bestimmte Probleme lösen zu können (König, 2015). Im engeren Sinn beschreibt Kompetenz nach Weinert zunächst ausschließlich kognitive Aspekte (Baumert & Kunter, 2011a). In einem weiteren Verständnis wird der Begriff „professionelle Handlungskompetenz“ durch die Übertragung des Kompetenzgedankens auf die Bewältigung beruflicher Anforderungen geprägt und schließt neben kognitiven Aspekten weitere Merkmale wie beispielsweise motivationale und soziale Bereitschaft und Fähigkeiten mit ein (Baumert & Kunter, 2011a). (Handlungs-) Kompetenz wird bezeichnet als

„persönliche Voraussetzung zur erfolgreichen Bewältigung spezifischer situationaler Anforderungen“ (Weinert, 2001, S. 27f.).

In zahlreichen Projekten (u.a. COACTIV-Studie Baumert & Kunter, 2011a; / FALKO-Studie Krauss et al., 2017b) wird auf wesentliche Aspekte des durch Weinert geprägten Kompetenzbegriffs Bezug genommen. Relevant scheint hier, dass Kompetenz in diesem Verständnis vermittelt und erworben werden kann, da sie im engeren Sinne kognitive Aspekte umfasst (Baumert & Kunter, 2011a). Damit ist angesprochen, dass es sich bei der Kompetenz um einen grundsätzlich *veränderbaren* Faktor handelt, der zudem in unterschiedlichen Situationen zum Tragen kommt (Handlungskompetenz) und somit kontextspezifisch ist (König, 2015). Kompetenz im engeren Sinne schließt an den Begriff des *Professionswissens* an, das auf das theoretische, deklarative, prozedurale und konditionale Wissen fokussiert, das „als *notwendige Voraussetzung für professionelles Handeln* im Unterricht zu verstehen [ist]“ (Krauss et al., 2017b, S. 38). Professionelle Kompetenz im weiteren Sinne umfasst also mehr als kognitive Aspekte und schließt das *Professionswissen* als Teil des gesamten Kompetenzkonstrukts mit ein. In Anlehnung an Weinert (2001) spielen neben kognitiven Aspekten (Professionswissen) die erwähnten affektiven Komponenten „Überzeugungen/Werthaltungen“, „Motivationale Orientierung“, „Selbstregulative Fähigkeiten“ eine entscheidende Rolle (Baumert & Kunter, 2006, S. 482).

Bei den Wissensbereichen wurden insbesondere die drei allgemein akzeptierten, auf Shulman (1986; 1987) zurückgehenden Bereiche *pädagogisches Wissen*, *Fachwissen* und *fachdidaktisches Wissen* intensiv beforscht und in zahlreichen Projekten zum fachbezogenen Professionswissen als eigene Wissenskategorien definiert (u.a. Baumert & Kunter, 2011a; Krauss et al., 2017b; Schumacher, 2016; Stancel-Piatak, Buchholtz & Schwippert, 2013). Bei dieser Trias handelt es sich um eine analytische Differenzierung von Wissenskomponenten, über die Lehrpersonen beim Unterrichten verfügen müssen (König et al., 2018). Eine genauere Beschreibung der Komponenten *Fachwissen* und *fachdidaktisches Wissen/Können* wird später in dieser Arbeit folgen (Kap. 3.3.1).

Wichtig scheint auch, dass kognitive Fähigkeiten als *Wissen* und *Können* definiert sind, womit der Handlungsbezug als konstituierend für kognitive Fähigkeiten als Teilbereich professioneller Kompetenz gilt (Seifert, 2015). Die Bedeutung des Handlungsbezugs (Können) wird auch im Begriff *Professionalität* aufgenommen. Professionalität wird im Gegensatz zum Professionswissen nicht während der Ausbildung erworben, sondern entwickelt sich in der Praxis im Umgang mit sich stellenden Anforderungen und unter dem Handlungsdruck des Schulalltags (Tenorth, 2006). Damit rückt der Begriff in die Nähe des kontextbezogenen Kompetenzbegriffs.

Entsprechend dem dargelegten Forschungsanliegen wird im Folgenden auf das *Professionswissen* fokussiert. Zusammenfassend zeichnen sich kognitive Fähigkeiten (Wissen und Können) als Teil professioneller Handlungskompetenz nach Seifert (2015) durch folgende vier Eigenschaften aus:

- 1) Sie liegen als Disposition zu Grunde und sind nicht direkt beobachtbar. Entsprechend müssen sie operationalisiert werden.
- 2) Sie sind grundsätzlich erlernbar und lassen sich durch pädagogische Maßnahmen beeinflussen.
- 3) Sie werden in spezifischen Situationen erworben und kommen in spezifischen Situationen zur Anwendung.
- 4) Sie haben einen Handlungsbezug und bilden sich damit im erfolgreichen Handeln in konkreten Situationen ab.

Kognitive Wissen und kontextbezogene Kompetenz als Kontinuum

Der angesprochene situationsspezifische Handlungsbezug führt zunächst zu einer konzeptionellen Unschärfe und zur Frage, was in Tests zum Professionswissen *genau* gemessen wird (Krauss et al., 2017b). Unklar ist, ob es sich um Kompetenzen im Sinne von dispositionalem, kognitivem Wissen handelt oder ob Kompetenzen im Sinne von Performanz erst durch die Anwendung in einer spezifischen Situation erfasst werden. Da Kompetenz u.a. als kognitive Fähigkeit und Fertigkeit verstanden wird, um bestimmte Probleme zu lösen, müssen zumindest analytisch und modellhaft „die einer Messung zugänglichen Dispositionen vom eigentlichen Vollzug der Problembewältigung in realen Anforderungssituationen“ unterschieden werden (König, 2015, S. 305). Um diesen Bruch zu überwinden, schlagen Blömeke, Gustafsson und Shavelson (2015a, S. 7) vor, Kompetenz als „Kontinuum“ zu begreifen. In diesem Verständnis ist die situationsspezifische Kompetenz als holistischer Prozess zu verstehen. Kontextbezogene Kompetenz mediert zwischen im Labor gemessenen Dispositionen und im Unterricht zu beobachtendem Verhalten (Performanz) und erfordert Wahrnehmungen, Interpretationen und Entscheidungen in domänenspezifischen Situationen (Krauss et al., 2017b). In diesem Kompetenzmodell (Abbildung 7) ist das Professionswissen (*Cognition*) im Feld der Disposition verortet, wird unter Verwendung affektiver Komponenten in der spezifischen Situation wirksam (*Situation-specific skills*) und schließlich als Leistung sichtbar (*Performance*). Der Bezug zum Weinert'schen Kompetenzbegriff wird auch in diesem Modell deutlich, da die Beteiligung kognitiver und affektiver Dispositionen an situationsspezifischen Lösungsprozessen aufgezeigt wird. Mit diesem Performanz-basierten Ansatz wird anstelle eines einseitigen Fokus auf das

Wissen von Lehrpersonen aber eine stärkere Verknüpfung von Wissen und Handeln verfolgt (König, 2015).

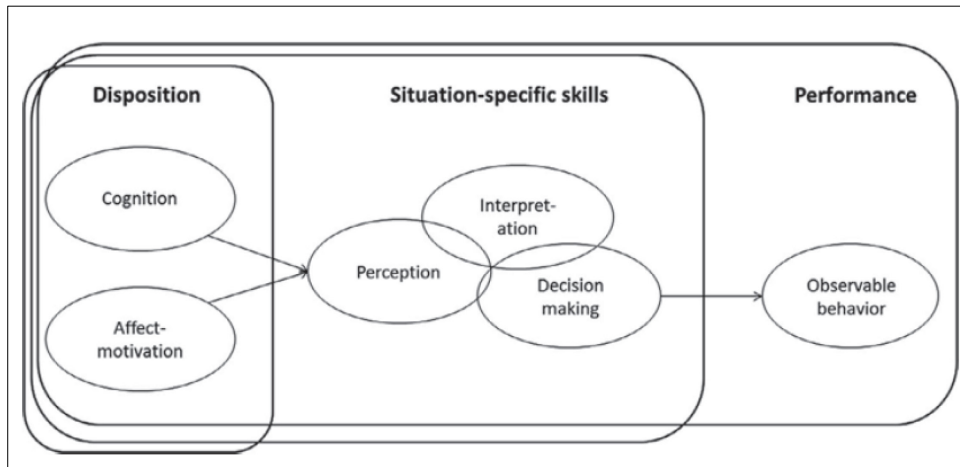


Abbildung 7. Kompetenz als Kontinuum (Blömeke et al., 2015a, S. 7)

Mit diesem *kontextbezogenen* Kompetenzbegriff gehen messmethodische Erweiterungen von reinen Papier-Bleistift Tests hin zu videobasierten Tests einher, bei denen Ausschnitte von Unterrichtssituationen als Stimuli für anschließende Fragen verwendet werden (König, 2015). Kompetenz bzw. Performanz wird hier näher an das tatsächliche Unterrichtsgeschehen gerückt. Letztlich bleibt der Kompetenzbegriff aber auch hier dem Paradigma der als *mental* gedachten Kompetenz verpflichtet (Neuweg, 2015, S. 379). Es zeigt sich nämlich, dass zur Absicherung von Soll-Antworten auch in (Video)-Vignetten basierten Studien meist Normen von Expertinnen und Experten beigezogen werden und dass die Annahme, dass mental vorliegende Kompetenzen durch den Prozess der Transformation in einen äußeren Vorgang der Performanz überführt werden, beibehalten wird (Neuweg, 2015).

Das Verhältnis von Wissen und Handeln wird im beschriebenen Kompetenzbegriff plausibel dargelegt. Ob diese „Annahme einer sich aus dem Theoriewissen durch Prozeduralisierung und Integration mehr oder weniger bruchlos entwickelnden berufspraktischen Kompetenz überhaupt richtig ist, und der Aufbau von Expertise als Prozess zunehmender Nutzung theoretischen Wissens im Laufe der Berufsbiografie verstanden werden kann“, bleibt gemäß Neuweg (2015, S. 379) allerdings offen. Fragen dazu, welches Wissen überhaupt wirksam ist (Brühwiler, Hollenstein, Affolter, Biedermann & Oser, 2017) und wie kognitive Wissensbestände im Lehrberuf in praktisches Handeln überführt werden (Scherf, 2013), bleiben weiterhin offen und bedürfen der empirischen Klärung. Oser (1997) stellt grundsätzlich fest, dass Wissen und Einsicht niemals Handeln garantieren kann, und „dass vielwissende Lehrpersonen nicht unbedingt gute oder erfolgreiche Lehrpersonen sind“ (S. 27).

Trotz offener Fragen ist der dargestellte Kompetenzbegriff als Kontinuum als Denkanstoß dennoch zentral. Eine Annäherung an Messverfahren, die stärker an reale Performanzsituationen heranrücken, ist in diesem Zusammenhang erstrebenswert und wird die Diskussion im Rahmen der Kompetenzmessung zur Lehrerinnen- und Lehrerprofessionalität weiterhin beeinflussen (Neuweg, 2015).

3.1.2 Verortung von Professionalität und Professionswissen in der kognitionspsychologischen Expertiseforschung

Wie in der Einleitung angedeutet, stellt der kognitionspsychologische Expertiseansatz im Rahmen der Forschung zum Professionswissen einen wichtigen Bezugspunkt für zahlreiche Studien dar. Die Begriffe „Expertin bzw. Experte“ und „Expertise“ sind aus der wissenschaftlichen Diskussion nicht mehr wegzudenken (Krauss & Bruckmaier, 2014, S. 243). Der Zusammenhang zwischen dem Professions- und dem Expertiseansatz besteht darin, dass sie im wesentlichen dieselben Phänomene erfassen (z.B. professionelles Wissen und Können von Lehrpersonen) (Krauss & Bruckmaier, 2014). Im Kern geht es in beiden Ansätzen darum, berufliche Aufgaben zu meistern, „die ein hohes Maß an komplexem, fachspezifischem Wissen erfordern, welches nur auf Basis langjähriger Ausbildung und Berufserfahrung zu erwerben ist“ (Reinisch, 2009, S. 38). Die deutliche Verbindung zeigt sich auch darin, dass sowohl der Expertiseansatz als auch Ansätze zur Messung von Professionswissen dem Forschungszweig der Kognitionsforschung zugeordnet werden und dass der Begriff des Expertinnen- und Expertenwissens mitunter mit dem Begriff der Professionalität bzw. dem Professionswissen gleichgesetzt wird (Krauss & Bruckmaier, 2014; Reinisch, 2009).

Expertise wird oft mit dem Aspekt der „Höchstleistung“ verbunden, wobei Baumert & Kunter (2011a, S. 34) die Leistungsbezogenheit relativieren und in Zusammenhang mit dem Wissen stellen, das ja üblicherweise bei Berufseinsteigerinnen und Berufseinsteigern (= Novizen) zumindest *geringer* sein müsste. Die Betonung liegt also eher auf graduellen Unterschieden und weniger auf „konstant hervorragenden Leistungen“ zur Lösung von Problemen (Krauss & Bruckmaier, 2014, S. 245). Damit wird der *leistungsorientierte* Expertisebegriff angesprochen, der sich vom *kompetenz- und wissensorientierten* Ansatz unterscheidet. Während im leistungsorientierten Ansatz nach Wissens- und Kompetenzunterschieden zwischen Expertinnen bzw. Experten und Berufsanfängerinnen bzw. Berufsanfängern gesucht wird, nähert man sich im wissensorientierten Ansatz eher theoretisch durch genaue Beschreibungen von Anforderungen (Krauss & Bruckmaier, 2014). Bereits bei Bromme (1992), der die Diskussion des Expertiseansatzes im Rahmen der Forschung zum professionellen Wissen von Lehrpersonen wesentlich geprägt hat, wird auf diese Doppeldeutigkeit des Expertinnen- und Expertenbegriffs

hingewiesen. Im Folgenden werden Überlegungen zu beiden Ansätzen skizziert, und es soll aufgezeigt werden, inwiefern insbesondere der kompetenz- und wissensorientierte Ansatz in aktuellen Modellen zum Professionswissen Eingang gefunden hat.

Leistungsorientierter Ansatz: Bei der leistungsorientierten Expertiseforschung geht es grundsätzlich darum, welches Fachwissen und welches Denken bei Expertinnen und Experten im Vergleich zu Novizinnen bzw. Novizen ausgewiesen werden kann. Der Blickwinkel ist dabei gemäß Zlatkin-Troitschanskaia & Kuhn (2010, S. 26) vor allem auf „kognitive Adaptionsprozesse“ gerichtet. Belege für die Anwendbarkeit des Experten-Novizen-Paradigmas für den Lehrberuf sind vor allem in den Arbeiten von Berliner (2001) zu finden (Baumert & Kunter, 2006; Krauss & Bruckmaier, 2014). Die Unterschiede zwischen Novizinnen bzw. Novizen und Expertinnen bzw. Experten legen die Annahme einer Entwicklung der Lehrerexpertise nahe, die sich als Abfolge von Entwicklungsstufen beschreiben lässt (König, 2010). Gemäß Berliner (2001) kann diese Entwicklung in fünf Stadien beschrieben werden, die in Abhängigkeit zu Ausbildungsgraden und Berufserfahrung stehen (Tabelle 1).

Tabelle 1: Adaptierte Darstellung der Entwicklungsstadien Lehrerexpertise nach Berliner (2001), abgedruckt in König (2010, S. 53)

1 Novize (<i>novice</i>)	Lehrerausbildung
2 Fortgeschrittener Anfänger (<i>advanced beginner</i>)	
3 Kompetent Handelnder (<i>competent performer</i>)	Berufseinstieg
4 Profilierter (<i>proficiency</i>)	Ab etwa 5. Berufsjahr
5 Experte (<i>expertise</i>)	

Baumert und Kunter (2006) fassen Berliners Ergebnisse dahingehend zusammen, dass es qualitative Unterschiede des Unterrichtens zwischen Novizinnen bzw. Novizen und Expertinnen bzw. Experten gibt. Gerade in kritischen didaktischen Situationen scheinen Expertinnen und Experten einen Vorteil in der Wahrnehmung zu haben, da ihre Kompetenz stärker „in Form von generellen Skripts organisiert ist“ (Baumert & Kunter 2006, S. 506). Während Expertenlehrpersonen auf die Arbeitsaktivitäten der Schülerinnen und Schüler fokussieren und Unterrichtssituationen schneller, genauer und ganzheitlicher wahrnehmen (Kaiser & Busse, 2015), richten Berufsanfängerinnen und Berufsanfänger die Aufmerksamkeit eher auf irrelevante Einzelheiten wie die Raumausstattung (Krauss & Bruckmaier, 2014).

Die Annahme, dass berufliche Entwicklungsprozesse im Verlaufe der Ausbildung geprägt werden und sich während der ganzen beruflichen Karriere fortsetzen, liegt auch dem Professionsverständnis der COACTIV-Studie zugrunde (Baumert & Kunter, 2011a). Damit wird der entwicklungsbezogene Aspekt professioneller Kompetenz beachtet. Dahinter steht die theoretische Annahme, dass im Rahmen der Ausbildung die Grundsteine professioneller Kompetenz gelegt

werden, die sich in den folgenden Berufsjahren in der praktischen Arbeit entwickelt (Baumert & Kunter, 2011a; Kunter, Kleickmann, Klusmann & Richter, 2011b; Terhart, 2005).

Beim leistungsorientierten Paradigma erweisen sich allerdings insbesondere zwei Aspekte als kritisch. Zum einen bleibt unklar, nach welchen Kriterien Expertenlehrpersonen (posteriori) identifiziert werden sollen (Krauss & Bruckmaier, 2014), und zum anderen ist der Zusammenhang zwischen der Anzahl Berufsjahre und der Unterrichtsqualität fraglich. Analysen der COACTIV-Studie zeigen zum Beispiel keinen positiven Zusammenhang zwischen fachdidaktischem Wissen, Fachwissen und Berufserfahrung (ebd.). Dennoch leistet dieser Ansatz auch in aktuellen Professionswissenstests (z.B. FALKO-Studie) einen Beitrag, indem Leistungsvarianzen durch den Einsatz in Kontraststichproben gemessen und damit Abgrenzungen zum Laienwissen festgestellt werden (Krauss et al., 2017b). Zudem können durch den Einsatz von Tests in verschiedenen Ausbildungsphasen Aussagen über potenzielle Erwerbsphasen bestimmter Wissensbestände während der Professionalisierung abgeleitet werden (ebd.).

Kompetenztheoretischer, wissensorientierter Ansatz: Der kompetenztheoretische Bestimmungsansatz baut auf der Expertiseforschung auf und definiert Kompetenzbereiche und Wissensdimensionen, die für die Bewältigung von Aufgaben im Lehrberuf wichtig sind (Bellenberg & Korte, 2019). Grundlegung für diesen wissensorientierten Ansatz sind demnach die Formulierungen von geteilten, professionsspezifischen *Anforderungen* bzw. *berufsspezifischen Kompetenzen* (Krauss & Bruckmaier, 2014). Diesem Ansatz folgend, werden diejenigen als Expertinnen und Experten bezeichnet, die eine spezialisierte, komplexe Aufgabe (z.B. das Unterrichten) erfolgreich bewältigen (Krauss & Bruckmaier, 2014). Die für den Lehrberuf erforderlichen Kompetenzen stehen dabei im Zusammenhang mit der Qualität von Unterricht und letztlich dem Lernzuwachs der Schülerinnen und Schüler (Bellenberg & Korte, 2019).

Die Arbeiten von Shulman (1987) und Bromme (1992) waren in diesem Paradigma richtungsweisend, weil damit Ordnungssysteme bzw. strukturbildende Wissenstaxonomien entwickelt wurden, die das professionelle Wissen für erfolgreiches Unterrichten beschreiben (Baumert & Kunter, 2011a; Krauss & Bruckmaier, 2014). Bromme (1992) ging es zentral darum, die Besonderheiten des professionellen Wissens empirisch und theoretisch zu fassen und den Begriff der Anforderungen zu präzisieren. Die einflussreichste Wissenstaxonomie, die u.a. im Modell der COACTIV-Studie (Kap. 3.2.2) Eingang gefunden hat, ist die Wissenstaxonomie von Shulman (Kap. 3.2.1). Er hat verschiedene Kerndimensionen bestimmt, die das Professionswissen von Lehrpersonen prägen (Baumert & Kunter, 2011a; Krauss & Bruckmaier, 2014). Diese Wissensdimensionen müssen kognitiv integriert werden, sodass sie eine Basis für beruflich relevantes Handeln darstellen (Blömeke, 2013). Als Expertise werden in diesem Verständnis vor

allem durch die *Ausbildung* veränderbare kognitive Kompetenzen (Professionswissen) und professionelle Überzeugungen verstanden (Krauss & Bruckmaier, 2014). Mit Bezug auf Bromme (2001) lässt sich zusammenfassen, dass insbesondere das domänenspezifische deklarative und prozedurale Wissen zum Können der Expertinnen und Experten beiträgt (König, 2010). Es geht in diesem Ansatz also nicht mehr um individuelle Bedingungen für Höchstleistungen, sondern um die „von den Mitgliedern einer Profession *geteilten*, besonderen berufsspezifischen Kompetenzen“ (Krauss & Bruckmaier, 2014, S. 251). In praktisch allen aktuellen Studien und Modellen zur Kompetenzmessung bei Lehrpersonen hat dieser kompetenztheoretische – auf das kognitive Wissen fokussierte – Expertiseansatz ganz direkt Eingang gefunden. Er war auch für die Entwicklung des vorliegenden Modells (Kap. 4) zentral.

3.2 Kompetenzmodelle und Typologien

Im Folgenden wird auf zwei Modelle Bezug genommen, die grundlegend sind für das in der vorliegenden Arbeit entwickelte Modell (Kap. 4). Es sind dies die Typologie nach Shulman und das COACTIV-Modell. Auch konzeptionelle Überlegungen der TEDS-Studien wurden für dieses Vorhaben übernommen. Sie betreffen jedoch den Aspekt der kognitiven Anforderungen von Wissen. Darauf wird in Kap. 3.3.3.2 eingegangen.

Im zeitlichen Ablauf steht die Typologie von Shulman (1986) am Anfang. Sie wurde später von Bromme (1992) und der Michigan-Gruppe (Hill, Ball Loewenberg & Schilling, 2008) aufgegriffen. Etwa zeitgleich mit der Testentwicklung der Michigan-Gruppe entstanden im Rahmen der MT21 und TEDS-M-Studie (Blömeke et al., 2008) sowie der COACTIV-Studie (Kunter et al., 2011a) im deutschsprachigen Raum Instrumente zur Erfassung des Professionswissens im Unterrichtsfach Mathematik. Eine Ausweitung auf andere Fächer erfolgte im Projekt FALKO (Krauss et al., 2017a) und mit der TEDS-LT-Studie (Blömeke, 2011; Blömeke et al., 2013a). Die Projektgruppen (COACTIV, TEDS-Studien, Michigan-Studie) arbeiteten weitgehend unabhängig voneinander und hatten unterschiedliche Zielgruppen im Blick (COACTIV-Studie: Sekundarlehrpersonen / MT21- TEDS-Studie: Lehramtsstudierende und Referendare / Michigan-Studie: Grundschullehrpersonen) (Krauss et al., 2017b). Die auf Shulman zurückgehende Unterscheidung von *fachdidaktischem Wissen* und *Fachwissen* sowie der Kompetenzbegriff von Weinert finden sich jedoch in nahezu allen Modellen wieder (ebd.). Insbesondere nimmt das COACTIV-Modell Bezug darauf. Das Modell von Shulman diente – teilweise in adaptierter domänenspezifischer Form – als Grundlage für die Entwicklung unzähliger Testinstrumenten (u.a. Bremerich-Vos, Buchholtz & König, 2016; Corvacho del Toro, 2013; Rutsch, Vogel, Seidenfuß, Dörfler & Rehm, 2018c; Schumacher, 2016; Wiprächtiger-Geppert et al., 2015).

3.2.1 Typologie professionellen Wissens nach Shulman

Den konzeptionellen Anstoß für die Lehrerkognitionsforschung gab Shulman (1986; 1987) mit seiner Modellierung des Wissens von Lehrpersonen (Reinisch, 2009). Ausgangslage für das Modell war die Annahme, dass Wissen und Können als Kern professioneller Kompetenz die wichtigste proximale Erklärungskomponente für den Lernerfolg von Schülerinnen und Schülern sind (Corvacho del Toro & Thomé, 2013). Einen entscheidenden Impuls für weitere Typologisierungen gab Shulmans Bestreben, die bis dahin getrennt gedachten Komponenten von „Content“ bzw. „Subject Matter“ (Fachwissen) und „Pedagogy“ (pädagogisches Wissen) in einen Zusammenhang zu bringen (Shulman, 1986, S. 6). Dies drückt sich in der folgenden Feststellung aus: „But no one focused on the subject matter content itself. No one asked how subject matter was transformed from the knowledge of the teacher into the content of instruction“ (ebd., S. 6). Aus der Einsicht, das fachdidaktische Wissen (*Pedagogical Content Knowledge*) als verbindende Kategorie zwischen pädagogischem Wissen und inhaltlichem Wissen aufzunehmen, entwickelte sich die bis heute meist gebräuchliche Trias von *Content Knowledge*, *Pedagogical Content Knowledge* sowie *General Pedagogical Knowledge* (Schumacher, 2016). Das fachdidaktische Wissen wird dabei als eine Art „Amalgam“ zwischen pädagogischem und fachlichem Wissen betrachtet (Shulman (1987, S. 8).

Shulman (1987, S. 8) formulierte zunächst sieben Kategorien, die für das Unterrichten als konstituierend erscheinen: (1) *Content Knowledge*, (2) *Pedagogical Content Knowledge*, (3) *General Pedagogical Knowledge*, (4) *Curriculum Knowledge*, (5) *Knowledge of Learners and their Characteristics*, (6) *Knowledge of Educational Contexts*, (7) *Knowledge of Educational Ends* (1987, S. 8). Von diesen Kategorien wurden in späteren Modellen aber meist nur die ersten drei bzw. vier berücksichtigt, weshalb im Folgenden auf diese fokussiert werden soll.

1) Mit *Content Knowledge* (CK) wird das inhaltliche Fachwissen bezeichnet. Damit sind Umfang und Organisation des inhaltlichen Wissens einer Lehrperson bezüglich eines bestimmten Faches gemeint (Shulman, 1987). Darüber hinaus müssen Lehrpersonen in der Lage sein, verschiedene Erklärungen zu einem Phänomen geben zu können und schließlich auch auf Fragen zu antworten, die über das „wie“ hinausgehen. „The teacher needs not only to understand that something is so, the teacher must further understand why it is so, on what grounds its warrant can be asserted“ (Shulman, 1986, S. 9). In den Worten von Kämper-van den Boogaart (2010) geht es hier um die Auswahl von Sachwissen, aber auch darum, Sachverhalte zu erklären und in einem größeren Zusammenhang richtig einzuordnen.

- 2) Das bei Shulman als *Pedagogical Content Knowledge* (PCK) bezeichnete Wissen entspricht gemäß Kämper-van den Boogaart (2010) am ehesten dem fachdidaktischen Wissen. Es geht hierbei darum, das Wissen in einem bestimmten Fach und in variablen Situationen passend zu unterrichten. Die Lehrperson muss Kenntnisse darüber haben, mit welchen Beispielen, Bildern und Analogien sie Sachverhalte für andere am besten verständlich machen kann (Shulman, 1987). Damit dies gelingt, muss der Lehrperson bekannt sein, welche Konzepte und Präkonzepte Schülerinnen und Schüler besitzen und was sie z.B. aufgrund ihres Alters oder aufgrund ihrer Herkunft mitbringen (Shulman, 1987). Nach Kämper-van den Boogaart (2010, S. 107) spielt im Zusammenhang mit dem *Pedagogical Content Knowledge* auch das „Fallwissen“ einer Lehrperson eine wichtige Rolle. Gemeint ist z.B. das Wissen über prototypische Fälle von Fehlkonzepten der Schülerinnen und Schüler (Shulman, 2004). Weiter spricht Shulman (1986) davon, dass Lehrpersonen über ein breites Repertoire an verständnisfördernden Repräsentationsformaten für das Fachwissen (Beispiele, Erklärungen, Demonstrationen usw.) verfügen und dass sie Bescheid wissen über die lernerleichternden oder -erschwerenden stoffbezogenen Kognitionen bestimmter Gruppen von Lernenden.
- 3) Das *General Pedagogical Knowledge* (PK) wird als allgemeines pädagogisches Wissen bezeichnet. Nach Shulman (1987) kann es verstanden werden als „knowledge, with special reference to those broad principles and strategies of classroom management and organization that appear to transcend subject matter“ (S. 8). Dieses Wissen ist nicht fachbezogen, sondern beinhaltet Kenntnisse über Lernende und ihre charakteristischen Eigenschaften, über Kontextfaktoren von Bildung, über Prinzipien der Klassenführung usw. (Lindmeier, 2011). Operationalisierungen zu diesem Aspekt finden sich etwa in der Übersicht von König et al. (2018). Hier werden auf der Grundlage verschiedener Studien Kategorien herausgearbeitet, die sich auf das Wissen über Voraussetzungen und Lernprozesse der Schülerinnen und Schüler, auf den Umgang mit dem sozialen Gefüge der Klasse, auf das methodische Repertoire sowie auf Fragen zur Leistungsbewertung beziehen.
- 4) Als weitere Kategorie definiert Shulman (1987, S. 10) das *Curricular Knowledge*, das auch im Michigan-Modell von Hill et al. (2008) Berücksichtigung findet. Zentral geht es um die Wahl angemessener Materialien, die für das Vermitteln von Inhalten verwendet werden. In diesem Zusammenhang weist Shulman (1987) darauf hin, dass das aktuell vermittelte Wissen mit dem parallel oder vorgängig sowie nachfolgend erworbenen Wissen in Verbindung gebracht werden sollte.

Die Typologisierung von Shulman hat im Rahmen der Lehrpersonenkognitionsforschung einen entscheidenden Beitrag geleistet, indem das professionelle Wissen, über das Lehrpersonen bei der Bewältigung berufsbezogener Anforderungen – und speziell beim Unterrichten – verfügen müssen, strukturell operationalisierbar gemacht wurde (König et al., 2018).

Allerdings gibt es auch Kritik am Modell, die bei Jandl (2016) vor allem mit Bezug auf das Review von Depaepe, Verschaffel und Klechtermans (2013) geäußert wird und sich im Wesentlichen „auf den Mangel an theoretischer und empirischer Grundlegung der Existenz von PCK als einer klaren Kategorie des professionelles Wissens von Lehrperson“ (S. 28f.) bezieht. Als weitere Kritik wird angeführt, dass die Sicht Shulmans auf das PCK zu starr sei, dass der unterrichtliche Kontext zu wenig berücksichtigt werde und dass der Handlungsbezug fehle (Jandl, 2016). Trotz kritischer Einwände hat sich jedoch die vorgelegte Differenzierung (CK, PCK, PC) nach Shulman weitgehend durchgesetzt (Kämper-van den Boogaart, 2010). Sie ist für das Verständnis weiterer Modelle hilfreich und hat zumindest im Sinne einer Grundlegung für die Konzeption weiterer, empirisch zu prüfender Systematisierungen ihren berechtigten Platz.

3.2.2 Das Kompetenzmodell der COACTIV-Studie

Rekonzeptualisierung des Modells von Shulman

Das Modell von Shulman erfuhr zunächst durch die Michigan-Gruppe (Ball, Thames & Phelps, 2008; Hill et al., 2008) eine „Rekonzeptualisierung“ (Jandl, 2016, S. 30). Es wurde am Fach Mathematik konkretisiert bzw. über entsprechende Items operationalisiert und damit empirisch zugänglich gemacht (Hill et al., 2008). Das Michigan-Modell wurde auch im europäischen Raum zur Kenntnis genommen und diskutiert. Aufgrund seiner empirischen Grundlegung war es richtungsweisend für das COACTIV-Modell (Baumert & Kunter, 2011a; Lindmeier, 2011). Allerdings war eine Weiterentwicklung notwendig, da sich die komplexe Struktur des Michigan-Modells in seiner Differenziertheit empirisch nicht bestätigen ließ (Baumert et al., 2010). Die beiden Hauptdimensionen *Content Knowledge* (Fachwissen) und *Knowledge of Content and Students* (fachdidaktische Komponente) konnten jedoch mit wenigen Einschränkungen grundsätzlich abgebildet werden (Hill et al., 2008). Die theoretische Modellierung der Wissenskomponenten des Michigan-Modells (ebd.) weicht zwar deutlich von der Modellierung bei COACTIV ab (Baumert & Kunter, 2011a), doch trotz aller Abweichungen lag beiden Projektgruppen die Idee zugrunde, das Professionswissen auf Grundlage der theoretischen Überlegungen von Shulman empirisch zugänglich zu machen und die Konstrukte „Fachwissen“ und „fachdidaktisches Wissen/Können“ zu operationalisieren.

Die COACTIV-Studie

Ein vielzitiertes Kompetenzmodell von Baumert und Kunter (2006) wurde im Rahmen der COACTIV-Studie (Kunter et al., 2011a) verwendet. Es bildet den theoretischen Rahmen für zahlreiche weitere Projekte der kognitionsorientierten Professionalitätsforschung bei Lehrpersonen. Die COACTIV-Studie wurde als eine von mehreren Ergänzungen im Rahmen der Teilnahme Deutschlands an der PISA-Studie durchgeführt (Löwen, Baumert, Kunter, Krauss & Brunner, 2011). Ziel war es, einerseits das Professionswissen von Lehrpersonen und andererseits Aspekte des kognitiv aktivierenden Mathematikunterrichts sowie die Entwicklung mathematischer Kompetenz der Lernenden zu untersuchen. Im Gegensatz zu anderen Studien fokussierte die COACTIV-Studie auf bereits praktizierende Mathematiklehrpersonen im 9. und 10. Schuljahr, die bei PISA 2003/2004 in Deutschland teilgenommen hatten (Löwen et al., 2011). Die Besonderheit lag darin, dass nicht nur Merkmale und Fähigkeiten von Lehrpersonen gemessen wurden und dass zwischen verschiedenen Schulformen (Hauptschulen / Sekundarschule) verglichen werden konnte, sondern dass durch die enge Verzahnung mit PISA Eigenschaften von Lehrpersonen mit dem Lernzuwachs der Schülerinnen und Schüler sowie mit weiteren unterrichtlichen Zielkriterien in Beziehung gesetzt wurden (Krauss et al., 2017b). Eine weitere Besonderheit lag in der längsschnittlichen Erhebung von Daten.

Die Bedeutsamkeit des COACTIV-Modells zeigt sich darin, dass es mehrfach repliziert wurde. In der FALKO-Studie (Krauss et al., 2017a) wurde das COACTIV-Modell direkt zugrunde gelegt, d.h. das konzeptuelle Vorgehen für die Operationalisierung des Fachwissens und des fachdidaktischen Wissens wurde zwecks Vergleichbarkeit aus dem Modell abgeleitet (Krauss et al., 2017b). Entsprechend lag der Fokus auf den kognitiven Wissensbereichen professioneller Kompetenzen. Das COACTIV-Modell wurde für die Testentwicklung in den Fächern Deutsch, Englisch, Latein, Physik, Musik und Religion verwendet. Schließlich wurde die Konzeption, die bei COACTIV zugrunde lag, auch in der EKoL-Studie einbezogen und „unter einer prozessorientierten Perspektive auf die Professionalisierung in der Lehramtsausbildung erweitert“ (Rutsch, Rehm, Vogel & Seidenfuß, 2018a, S. 10).

Das Modell in der Übersicht

Das COACTIV-Modell baut auf den Strukturen Shulmans sowie auf deren Erweiterungen (z.B. durch Bromme) auf (Löwen et al., 2011). „Professionelle Kompetenz“, wie sie im COACTIV-Modell genannt wird, speist sich aus verschiedenen Forschungstraditionen, wobei die Betonung auf dem „Wissensbereich“ liegt – also auf kognitiven Merkmalen von Lehrpersonen, die in der Tradition der Expertiseforschung zentral sind (Baumert & Kunter, 2011a, S. 45). Dieser Ansatz

bildet eine Gegenposition zum Verständnis, wonach der Lehrberuf auf Talent und angeborenen Dispositionen beruht (ebd.). Mit der Berücksichtigung von Überzeugungen, Werthaltungen, motivationaler Orientierung und Selbstregulationsfähigkeit geht das Modell über die Auffassung von Expertise und einen rein auf Kognitionen bezogenen Begriff von Kompetenz hinaus (ebd.). Eine Erweiterung der Typologisierung nach Shulman findet durch den Einbezug von *Organisationswissen* und *Beratungswissen* statt (Schumacher, 2016, S. 33). Diese dem Professionswissen zugeordneten Bereiche grenzen sich vom Fachwissen und vom fachdidaktischen Wissen einerseits und vom pädagogischen Wissen andererseits ab und erfordern eigene Messinstrumente (Baumert & Kunter, 2011a). Trotz des Einbezugs dieser Dimensionen im Kompetenzmodell beschränkte sich die empirische Überprüfung auf die drei Wissensbereiche *Fachwissen*, *fachdidaktisches Wissen* und *pädagogisches Wissen*, da diese Trias eine unmittelbare Handlungsrelevanz für die Unterrichtsgestaltung aufweisen.

Im Modell wurden verschiedene Ebenen unterschieden, die sich in *vier Kompetenzfaktoren* (Professionswissen, Überzeugungen, Motivation und Selbstregulation) sowie in spezifische *Kompetenzbereiche* und deren *Kompetenzfacetten* gliedern (Abbildung 8).

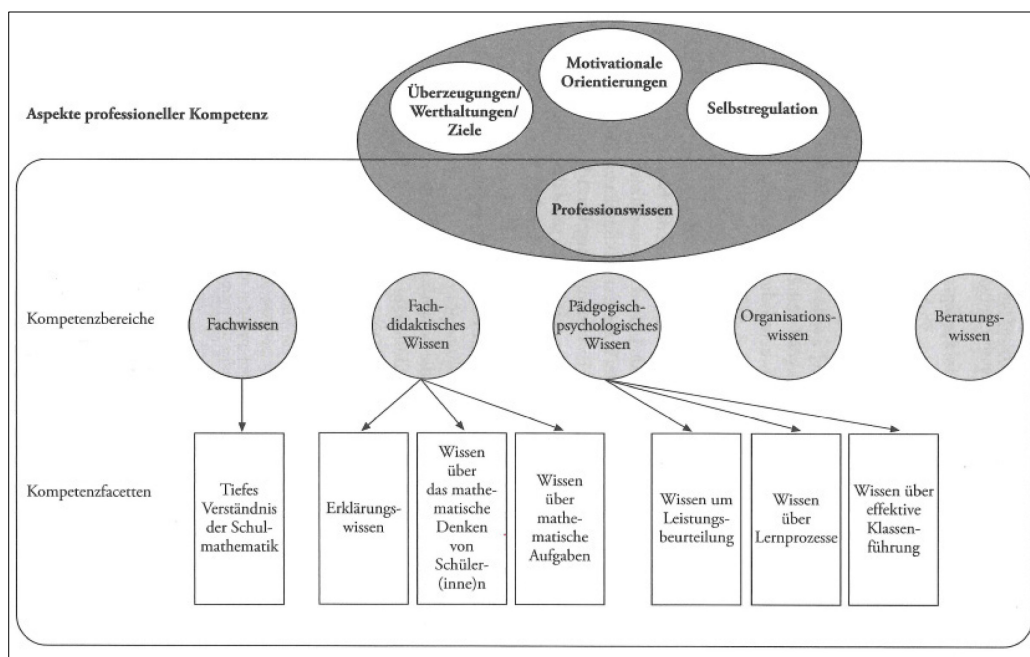


Abbildung 8. Modell professioneller Kompetenz bei COACTIV mit Spezifikationen für das Professionswissen (Baumert & Kunter, 2011a, S. 32)

Dem Forschungsanliegen der vorliegenden Studie folgend, wird der Fokus auf den Bereich des *Professionswissens* gelegt. Im Besonderen werden die Bereiche Fachwissen und fachdidaktisches Wissen/Können genauer beschrieben. Zuvor jedoch ein Wort zum *pädagogisch-psychologischen Wissen*, das als fachunabhängig beschrieben wird und im Ordnungsschema nach Shulman (1987) dem *General Pedagogical Knowledge* zuzuordnen wäre. Hierzu zählen

Facetten wie die Klassenführung, die Diagnostik bzw. das Wissen um Leistungsbeurteilung und das Wissen über Lernprozesse. Eine detaillierte Darstellung dieser Facetten findet sich bei Baumert & Kunter (2011a). Dieser Kompetenzbereich wurde auch in neueren Studien operationalisiert und empirisch erfasst. Allerdings bleibt die Frage offen, inwiefern dieser Bereich tatsächlich ohne fachliche Konkretisierung gedacht werden kann oder ob auch hier domänenspezifisches Wissen nötig ist (König et al., 2018). Diese Frage stellt sich deutlich im Bereich der Leistungsbeurteilung und Diagnostik oder beim Wissen über Lernprozesse.

Wie bereits im Modell der Michigan-Gruppe bzw. bei Shulman wird im COACTIV-Modell zwischen Fachwissen und fachdidaktischem Wissen unterschieden.

Fachwissen

Im Unterschied zur Michigan-Gruppe, die zwei formale Wissensdimensionen unterscheidet (*Common Knowledge* und *Specialized Knowledge*) geht COACTIV (Baumert & Kunter, 2011a) von vier Formen mathematischen Wissens aus, „welche jeweils unterschiedliche Grade der stofflichen Durchdringung widerspiegeln“ (S. 37). Die Konzeptualisierung und Operationalisierung wird in Anlehnung an Shulman (1986) Charakterisierung des *content knowledge* vorgenommen und in folgenden vier Niveauebenen umschrieben (Krauss et al., 2011, S. 142):

Ebene 1: Mathematisches Alltagswissen, über das grundsätzlich alle Erwachsenen verfügen sollten.

Ebene 2: Beherrschung des Schulstoffs (etwa auf dem Niveau eines durchschnittlichen bis guten Schülers der jeweiligen Klassenstufe).

Ebene 3: Tieferes Verständnis der Fachinhalte des Curriculums der Sekundarstufe (z.B. auch Elementarmathematik vom höheren Standpunkt aus, wie sie an der Universität gelehrt wird).

Ebene 4: Reines Universitätswissen, das vom Curriculum der Schule losgelöst ist (z.B. Galoistheorie, Funktionsanalyse).

Damit wird angedeutet, dass das zu erreichende Niveau mit der Schulstufe, auf der Lehrpersonen später unterrichten, zusammenhängt, was in anderen Studien explizit formuliert wurde (z.B. MT21 Studie: Mathematik Sekundarstufe I und Sekundarstufe II, Schulmathematik vom höheren Standpunkt, Universitäre Mathematik) (Niermann, 2017). In der COACTIV-Studie wurde lediglich das Wissen auf Ebene 3 erhoben, was bedeutet, dass Lehrpersonen über qualitativ höheres Wissen verfügen müssen als den von ihnen unterrichtende Stoff (Krauss et al., 2011). „Dieses professionelle Fachwissen schließt die Beherrschung des Schulstoffs selbst mit ein;

aber weder solches Schulwissen, geschweige denn mathematisches Alltagswissen genügen, um die mathematischen Herausforderungen zu bewältigen, die sich Lehrkräften bei der Vorbereitung und Durchführung des Unterrichts stellen“ (Baumert & Kunter, 2011a, S. 37).

Fachdidaktisches Wissen

Vom fachlichen Wissen unterschieden wird das fachdidaktische Wissen, das allerdings mathematisches Wissen voraussetzt (Kunter et al., 2011a). Ganz allgemein wird damit das „Zugänglich machen mathematischer Inhalte für Schüler“ umschrieben (Krauss et al., 2011, S. 138). Diese Definition wurde auch in den parallel laufenden Michigan-Studien (Hill et al., 2004) bei Grundschullehrpersonen und bei TEDS-M (Blömeke et al., 2008) für Sekundarlehrpersonen angewendet (Krauss et al., 2011). Operationalisiert wurde es bei COACTIV über drei Facetten:

- 1) *Zugänglichmachen / Wissen über Erklärungen und Repräsentationen*: Dies umschreibt das Wissen über Repräsentationen und Erklärungen einschließlich des schnellen Erkennens von Fehlern (Baumert & Kunter, 2011a; Helmke, 2012). Lehrpersonen sollten in der Lage sein, Sachverhalte auf geeignete Weise zu erklären und Inhalte verständlich zu machen, was voraussetzt, dass sie über ein Repertoire an Erklärungsmöglichkeiten verfügen (Krauss et al., 2011).
- 2) *Schülerinnen und Schüler / Wissen über typische Fehler und Schwierigkeiten*: Dieses Wissen bezieht sich auf mathematische Vorstellungen sowie Fehlkonzepte bzw. typische Fehler von Schülerinnen und Schülern. Das Erkennen von Schwierigkeiten kann als didaktische Chance für Lernprozesse genutzt werden, was voraussetzt, dass Fehler erkannt, analysiert und konzeptuell eingeordnet werden (Krauss et al., 2011). Hier wird explizit auch die Diagnostik von Schülerinnen- und Schülerwissen sowie von Verständnisprozessen erwähnt (Baumert & Kunter, 2011a).
- 3) *Inhalte / Wissen über das multiple Lösungspotenzial von Aufgaben*: Gemeint ist das Wissen über das diagnostische und didaktische Potenzial sowie über die kognitiven Anforderungen und Wissensvoraussetzungen von Aufgaben (Krauss et al., 2011). Dazu gehören auch Vorstellungen zur *Passung* von Aufgaben (Helmke, 2012) sowie Kenntnisse über die didaktische Sequenzierung von Aufgaben sowie die langfristige curriculare Anforderung des Schulstoffs (Baumert & Kunter 2011, 37).

Die drei Kompetenzbereiche – das pädagogisch-psychologische Wissen, das Fachwissen sowie das fachdidaktische Wissen – haben sich als Kernkategorien des Professionswissens von Lehrpersonen etabliert (Pissarek & Schilcher, 2015; Rutsch, 2016). Krauss et al. (2008) untermauern

dies, wenn sie feststellen: „Es besteht kein Zweifel, dass allen dreien eine zentrale Bedeutung bei der professionellen Aufgabe der Lehrerinnen und Lehrer zukommt“ (S. 226). Die Beschreibung der drei Kompetenzbereiche bzw. deren Operationalisierung (auf Ebene von Kompetenzfacette) wurde in zahlreichen Studien (siehe oben) aufgenommen, ausdifferenziert und auf diverse Domänen übertragen (vgl. Übersicht zu Studien in verschiedenen Fächern in Krauss et al., 2017a). Auch die enge Verflechtung der beiden Bereiche Fachwissen und fachdidaktisches Wissen waren Gegenstand eines intensiven theoretischen Diskurses und verschiedener empirischer Untersuchungen (Helmke, 2012). Auf die Operationalisierung des Professionswissens und die Zusammenhänge von Fachwissen und fachdidaktischem Wissen wird im Folgenden genauer eingegangen, und zwar wird eine Annäherung an die Domäne *Lesen* versucht. Die vorgestellten Modelle sind dabei immer Referenz für weiterführende Überlegungen.

3.3 Operationalisierung des Professionswissens

Mit Bezug auf die bisherigen Darlegungen zum Professionswissen – mit Fokus auf das Fachwissen und das fachdidaktische Wissen – wird im Folgenden genauer umschrieben, wie die Konstrukte operationalisiert werden können. Dabei werden Überlegungen angestellt, wie der Transfer auf die Domäne des Lesens vollzogen werden kann. Neben dieser inhaltlichen Beschreibung folgen Reflexionen zur formalen Kategorisierung (Rutsch, 2016), d.h. zur mentalen Repräsentation von Wissen. Unter Einbezug messmethodischer Überlegungen wird versucht, die Konstrukte für das hier entwickelte Modell operationalisierbar zu machen und deren Relevanz hinsichtlich ihres Einflusses auf Schülerinnen- und Schülerleistungen aufzuzeigen.

3.3.1 Konzeptualisierung von „*Fachwissen*“, „*fachdidaktisches Wissen/Können*“ und „*diagnostisches Wissen*“

3.3.1.1 Fachwissen

Herausforderung zur Beschreibung des „Inhalts“ von Fachwissen (Content Knowledge – CK)
Nachfolgend werden Aspekte angesprochen, die hinsichtlich der Operationalisierung des Fachwissens einbezogen werden müssen. Als erstes stellt sich die Frage, auf welchem Niveau das fachliche Wissen in der Domäne *Lesen* umschrieben werden kann bzw. was der „Inhalt“ in diesem Fall überhaupt sein soll, da es nicht darum gehen kann, die Lesekompetenz von Lehrpersonen auf Primarstufenniveau zu erfassen. Unter Rückbezug auf bestehende, teilweise fachfremde Konzeptualisierungen müssen im vorliegenden Projekt Kategorien zum *leseerwerbsbezogenen Fachwissen* und zum (*diagnostischen*) *Wissen über Leseschwierigkeiten* gefunden werden. Die Schwierigkeit, solche Kategorien zu finden, zeigt sich hier deutlich. Dies hängt

damit zusammen, dass man bei der Konstruktion eines Erhebungsinstruments auf Probleme stößt, „wenn in einer Domäne noch relativ wenig über belegbare und belastbare Standards bekannt ist, also nur für wenige Kompetenzen geklärt ist, welche Kompetenzstufen und -dimensionen anzunehmen sind“ (Pissarek & Schilcher, 2015, S. 329). Hilfreich können zum einen Normen von Expertinnen und Experten und der Bezug auf gängige fachdidaktische Quellen sein, die sich auf empirisch haltbare Forschungsergebnisse stützen. Zum anderen erweist sich auch die Orientierung an Bildungsstandards und Lehrplänen als sinnvoll (ebd.).

Die Suche nach geeigneten Kategorien zum Fachwissen ist auch deshalb herausfordernd, weil bisher genannte Umschreibungen von Fachwissen allgemeiner Art waren (z.B. bei Shulman) oder sich auf den Fachbereich Mathematik bezogen. Eine Ausweitung auf „gering strukturierte Domänen“, wie die Fächer Deutsch und Englisch bei TEDS-LT (Blömeke, 2011) bezeichnet werden, fand erst später statt und wirft neue Fragen zur Operationalisierung dieses Fachwissens auf. Die geringere Strukturierung wird u.a. darin gesehen, dass das curriculare Angebot weniger strukturiert ist als jenes der Mathematik und dass sich die Entwicklung von Items mit „richtig“-„falsch“-Antworten von denjenigen im Fach Mathematik unterscheidet (ebd.).

Mit Blick auf Testinstrumente zur validen Erfassung kognitiver Wissensbestände im Bereich Lesen (Rutsch, 2016) bzw. Schreiben (Corvacho del Toro, 2013; Jagemann, 2016; Sturm, Schneider, Lindauer & Sommer, 2016; Wiprächtiger-Geppert et al., 2015) zeigen sich die besonderen Herausforderungen hinsichtlich der Operationalisierung deutlich. Lessing-Sattari und Wieser (2018) sehen die Problematik der Entwicklung valider Testinstrumente in diesem Bereich unter anderem darin, dass die Abgrenzung von fachlichem und fachdidaktischem Wissen nicht immer einfach ist und dass sich die Modellierungen dieser Wissensbereiche in den einzelnen Studien durchaus unterscheiden. „Aufgrund der Überschneidungen von psychologischer und fachdidaktischer Schreib- und Leseforschung sind solche Zuordnungsschwierigkeiten nicht verwunderlich“ (Lessing-Sattari & Wieser, 2018, S. 45). Ein Grund dafür liegt nach König et al. (2018) darin, dass zum einen z.B. im Fach Deutsch „Inhalt zugleich Medium der Lehre [ist], sodass hier die Unterscheidung von PCK und CK zumindest partiell anders als üblich zu fassen sein dürfte“ (S. 8) und zum anderen, dass „die Konturen der Sprachfächer im Vergleich mit Mathematik und den Naturwissenschaften recht diffus [sind]“ (ebd., S. 8). Phelps & Schilling (2004) sehen die Schwierigkeiten und Gründe für die anfängliche Vernachlässigung der Operationalisierung kognitiven Wissens im Bereich Lesen im Gegensatz zur Mathematik und zu naturwissenschaftlichen Fächern darin, dass Lesen kein eigenes Schulfach ist. Gemäß den Autoren scheint es unklar, wie das inhaltliche Wissen im Bereich Lesen überhaupt identifiziert wird. Weiter stellen Phelps & Schilling fest, dass viele

Lehrpersonen im Bereich Mathematik oder in naturwissenschaftlichen Fächern substanzielle Verständnismängel haben und dass der Mangel an inhaltlichem Wissen auch wahrgenommen wird. Im Gegensatz dazu sind die allermeisten Grundschullehrpersonen kompetente Leserinnen und Leser. Das Wissen um hinter dem Leseprozess stehende Strukturen wird folglich viel weniger wahrgenommen (ebd.).

Zur Operationalisierung von Fachwissen: Annäherung und Rückbezug auf andere Domänen

Der Blick auf die Operationalisierung in der Domäne Rechtschreiben, bei der – zumindest in gewissen Teilen – ähnliche theoretische Bezüge zum Schriftspracherwerb hergestellt werden, erweist sich als fruchtbar. Riegler und Wiprächtiger-Geppert (2016) operationalisieren das Fachwissen zum Orthographieerwerb unter Rückbezug auf die Definition der COACTIV-Studie im Sinne eines tieferen Verständnisses der Inhalte des Schreibunterrichts der Grundschule.

Das bedeutet, dass über das eigene Rechtschreibenkönnen hinaus erwartet wird, dass Lehrpersonen einerseits Schüleraufgaben aus dem Stand lösen können und andererseits in der Lage sind, die zugrunde liegenden Regularitäten angemessen, das heißt fachlich korrekt zu verbalisieren und fachsprachlich dargestellte Regularitäten rezeptiv zu verarbeiten. (Riegler & Wiprächtiger-Geppert, 2016, S. 205)

Einen praktisch identischen Bezug machen Corvacho del Toro und Thomé (2013) bzw. Corvacho del Toro (2013) in ihren Studien zu Wissensbereichen der Schriftsprache in den ersten Schuljahren aus linguistischer und psycholinguistischer Perspektive. Sie erfassen das Fachwissen in einer Facette und bezeichnen damit das tiefe Verständnis des deutschen Schriftsystems, des Gegenstands Orthographie und des Schriftspracherwerbs (Corvacho del Toro & Thomé, 2013). Die Dimension „Fachwissen“ wird in der Studie von Corvacho del Toro (2013) in zwei Teilen erfasst: Teil A befasst sich mit der Phonologie und Graphematik des Deutschen und wird meist über geschlossene Frageformate erhoben, Teil B beschäftigt sich mit dem Fachwissen über Phasen des Schriftspracherwerbs (Corvacho del Toro & Thomé, 2013). Aufgaben zu *Teil A* bestehen u.a. in der Analyse von Silbenstrukturen von Wörtern, im Bestimmen von lauttreuen Wörtern, von Graphemen, Phonemen, Morphemen, im Nennen der Anzahl der deutschen Vokale und in der Lautanalyse von Wörtern (Corvacho del Toro, 2013). Aufgaben zu *Teil B* beziehen sich auf die voralphabetischen, alphabetischen und orthographischen Phasen des Schriftspracherwerbs (ebd.).

Deutliche Parallelen der Erfassung des Fachwissens in Teil A lassen sich in der Studie von Riegler und Wiprächtiger-Geppert (2016, S. 205) finden, die das Fachwissen in den Wissensfacetten „Phonetisch-phonologisches Wissen“ und „Graphematisches Wissen“

operationalisieren (siehe auch Jagemann, 2016). Darin werden Aspekte zu Phonemanalyse, Graphem-Phonem-Korrespondenz und Regularitäten der deutschen Sprache abgefragt (Riegler & Wiprächtiger-Geppert, 2016). Eine gewinnbringende Erweiterung findet sich in der Studie zum Wissen von Lehrpersonen über Schreibprozesse und textlinguistisches Wissen von Sturm et al. (2016). Hier wird das Wissen über Merkmale schwacher Schreiber und Schreiberinnen zur Dimension des fachlichen Wissens gezählt. Übertragen auf die Domäne Lesen gehört zum Professionswissen entsprechend auch das Wissen über Merkmale leseschwacher Kinder.

Bezüge zum Fachwissen im Bereich Lesen aus dem angloamerikanischen Raum

Die Frage nach dem Inhalt von Leseunterricht auf Primarstufe und bezüglich Analogien zum Fach Mathematik stellen sich Phelps und Schilling (2004) „What might be the analog of this sort of content knowledge in reading?“ (S. 33). Ihre Antwort findet sich in der Operationalisierung zu Wissensbeständen im Bereich Erstleseunterricht. Ihre Fragestellung bezieht sich darauf: „What do reading teachers need to know about language, text, and reading processes, and in what ways, in order to help others learn to read?“ (ebd., S. 33). Die Autoren beantworten die Frage dahingehend, dass sie in ihrer Studie die zwei Gegenstandsbereiche *Comprehension (CMP)* und *Word analysis (WA)* unterscheiden. Dieses Wissen betrifft zusammenfassend das inhaltliche Wissen über Aspekte des Lesens im Kontext bestimmter Lehrsituationen.

Comprehension: In diesem Bereich werden u.a. Aspekte der Morphologie, des Vokabulars, Leseverstehensstrategien, Verständnis von Wörtern und Texten und Aspekte der Leseflüssigkeit zusammengefasst.

Word Analysis: Dieser Bereich beinhaltet Aspekte der phonologischen Bewusstheit, Phonem-Graphem-Korrespondenz, Betonungen von Wörtern und weitere Aspekte im Zusammenhang mit dem Dekodieren von Wörtern und ihrem schriftlichen und lautlichen Erscheinungsbild.

Betreffend Fachwissen (KC) wird davon ausgegangen, dass zum einen spezielles linguistisches Wissen über die Struktur der Sprache notwendig und zudem Kenntnisse über Leseprozesse entscheidend sind (Phelps & Schilling, 2004). Lehrpersonen müssen z.B. entsprechende Kenntnisse über Buchstaben-Laut-Beziehungen, über Prozesse beim Dekodieren von Wörtern sowie grundlegendes Wissen über das Textverständnis haben (ebd.). Es geht also um ein spezialisiertes inhaltliches Wissen über sprachliche Aspekte, das über das Wissen hinausgeht, über das Erwachsene typischerweise verfügen (ebd.).

Auch in der Studie von Moats & Foorman (2003) finden sich Hinweise zur Operationalisierbarkeit von lehrpersonenbezogenen Wissensgrundlagen im Bereich Lesen. Moats und Foorman

(2003) haben stufenabhängige Testformen entwickelt, die jeweils spezifisches lesebezogenes Wissen von Lehrpersonen auf der Stufe Kindergarten bis zweite Klasse (Form 1), auf der Stufe zweite und dritte Klasse (Form 2) und auf der Stufe vierte und fünfte Klasse (Form 3) erfassen: *Form 1:* Hier wurde sowohl über Multiple-Choice- als auch über offene Fragen Aspekte zur phonologischen Bewusstheit, Phonem-Graphem Korrespondenz, Definition von wichtigen Konzepten zur Leseentwicklung und zum Umgang mit Silben erfasst.

Form 2: Hier wurden über Multiple-Choice-Fragen Aspekte zur Identifikation von Phonemen, von Silben, von Wortstrukturen erfasst und zudem das Verständnis des Unterschiedes zwischen Hörverstehen und Leseverstehen erfragt.

Form 3: Bei dieser Form wurden Fragen im Multiple-Choice-Format sowie Fragen zur Interpretation von Schülerinnen- und Schülerprodukten im Bereich Lesen und Schreiben gestellt. Inhaltlich wurden Aspekte aus den Bereichen Orthographie, Phonologie, Morphologie sowie der Leseinstruktion und des Erkennens und Deutens von Lese- und Rechtschreibfehlern berücksichtigt.

Bezüge zum Fachwissen im Bereich Lesen und LRS aus dem deutschsprachigen Raum

Antworten auf die Operationalisierbarkeit im Bereich Lesen finden sich auch in der Studie von Schmidt und Schabmann (2016). In der Untersuchung wurde ein Wissenstest eingesetzt, mit dem das fachliche und fachdidaktische Grundwissen über LRS, das linguistische Grundwissen sowie das Wissen im Bereich Lesen und Leseentwicklung bei Referendarinnen und Referendaren erhoben wurde (ebd.). Fachwissen und fachdidaktisches Wissen wurden nicht in getrennten Skalen erfasst. Da die Entwicklung dieses Instruments etwa zeitgleich mit dem vorliegenden Projekt stattfand, konnten diese Items nicht übernommen werden. Trotz unabhängiger Operationalisierung zeigt sich eine gewisse Übereinstimmung der Items mit denjenigen aus der vorliegenden Arbeit, was im Zusammenhang mit der inhaltlichen Validität als erfreulich gewertet werden kann. Tabelle 2 zeigt die zehn verwendeten Items für die Skala Grundwissen zum Lesen/LRS. Alle Aufgaben wurden im MC-Format mit vier Antwortalternativen vorgegeben.

Tabelle 2: Ausformulierte Items und Lösungswahrscheinlichkeiten für die Skala Grundwissen zum Lesen/LRS aus (Schmidt & Schabmann, 2016)

LRS [Korrekte Antwort]	Lösungswahrscheinlichkeit
Pseudowörter sind... [leicht aussprechbare sinnfreie Buchstabenfolgen]	.37
Ab welchem Zeitpunkt können bereits Probleme beim Schriftspracherwerb auffallen? [Bereits drei bis vier Monate nach der Einschulung]	.59
Was sind Kennzeichen evidenzbasierter Förderung? [Programme sind wissenschaftlich empirisch überprüft]	.55
Worin kann eine niedrige Lesegenauigkeit bei unbekannten Wörtern begründet sein? [Kinder beherrschen Buchstaben-Laut-Beziehung nicht]	.61
Was ist ein bekanntes Verfahren zur Feststellung der Lese-Rechtschreibkompetenz? [Salzburger Lese-Rechtschreibtest (SLRT)]	.23
Was ist kein Risikofaktor für die Ausbildung von Leserechtschreibschwierigkeiten? [Linkshändigkeit]	.51
Unbehandelte Lese-Rechtschreibschwierigkeiten [..."wachsen" sich in der Regel nicht aus]	.76
Leseprobleme werden in der Mehrzahl der Studien häufiger [bei Jungen als bei Mädchen gefunden]	.59
Wobei handelt es sich um KEINE Standardaufgabe zur phonologischen Bewusstheit? [Buchstaben nachzeichnen]	.64
Ein Phonem ist die... [kleinste bedeutungsunterscheidende Einheit der Schriftsprache]	.32

3.3.1.2 Fachdidaktisches Wissen

Beschreibung und Herausforderungen zum fachdidaktischen Wissen (Pedagogical Content Knowledge – PCK)

Zur Beschreibung von fachdidaktischem Wissen liegt eine Fülle von Arbeiten aus den vergangenen dreißig Jahren vor, aufgrund deren versucht wurde, Gemeinsamkeiten zu identifizieren (König et al., 2018). Viele Konzeptionen von PCK berufen sich auf zwei Facetten: „Wissen über Lehrmethoden des Faches (einschließlich Wissen um fachliche Darstellung und Erklärungen) und Wissen über das Lernen und Verständnis von Schülerinnen und Schülern, einschließlich der Antizipation von Verständnisschwierigkeiten von Schülerinnen und Schülern im betreffenden Fach“ (ebd., S. 8). Wie in vielen ähnlichen Umschreibungen (u.a. Neuweg, 2014) wird hier der Bezug zum Konzept von PCK von Shulman (1987) deutlich. Über die bereits genannten Facetten hinaus beinhaltet das Wissen mit Bezug auf Shulman auch das Wissen über das Fachcurriculum und Kenntnisse darüber, welche Methoden und Medien in welcher Altersphase und bei bestimmten Inhalten angebracht sind (Neuweg, 2014; Seifert, 2015). Dieser Bereich umfasst demnach *alles* Wissen, das relevant ist, damit es in einem bestimmten Fach vermittelt werden kann (Seifert, 2015).

Des Weiteren scheint das fachdidaktische Wissen stärker als das Fachwissen an bestimmte unterrichtliche Kontexte gebunden zu sein und entsteht durch aktive Konstruktions-, Integrations- und Transformationsleistungen der Lehrpersonen in Handlungen im Unterricht (Neuweg, 2014). Von einem stärkeren Handlungsbezug gehen auch Pissarek und Schilcher (2017) bei der Operationalisierung des fachdidaktischen Wissens im Rahmen der FALKO-D-

Studie aus. Am Beispiel der Operationalisierung des fachdidaktischen Wissens zur Facette *Auswahl und Potenzial von Texten* wird aufgezeigt, dass das fachdidaktische Wissen über bloßes Fachwissen zu Textmustern hinaus geht, „so dass fachdidaktisches und fachliches Wissen stets zusammenwirken und hier im Sinne einer Kompetenzorientierung meist eine «Handlungskomponente» in den Items implementiert wurde“ (ebd., S. 79). In diesem Zusammenhang wird meistens auch der Begriff des *Könnens* in Abgrenzung zum *Wissen* eingeführt. Neuweg (2014) bringt dies auf den Punkt, wenn er feststellt, dass in diesen Handlungen ein komplexes Können zum Ausdruck komme, das regelmäßig als fachdidaktisches Wissen bezeichnet werde. Generell kann angenommen werden, dass fachdidaktisches Wissen stark prozeduralisierte Wissensanteile beinhaltet (Vogelsang & Reinhold, 2013).

Auch wenn hiermit erste Anhaltspunkte zur Struktur und zur mentalen Repräsentation des fachdidaktischen Wissens bzw. Könnens gewonnen werden können, stellt sich die Frage, wie sich diese auf andere Domänen übertragen lassen, da ja erste Ansätze auch hier meist auf den Bereich Mathematik und naturwissenschaftliche Fächer begrenzt waren (Rutsch, Seidenfuß, Vogel & Rehm, 2017). Vermutet wird, dass in der Konzeptualisierung von PCK in den bisher untersuchten Fächern auf der einen und in den sprachlichen Fächern auf der anderen Seite deutliche Differenzen bestehen (König et al., 2018).

Die Herausforderung stellt sich dadurch, dass sprachliche Fächer diffusere Konturen aufweisen als beispielsweise naturwissenschaftliche Unterrichtsgegenstände und dass oft zahlreiche konkurrierende Theorien bestehen (König et al., 2018). Fragen zur inhaltlichen Validität fachdidaktischen Wissens müssen geklärt werden. Es muss offengelegt werden, an welchen Normen man sich orientiert (Zimmermann & Peyer, 2016). So wurde etwa in der FALKO-Studie zur inhaltlichen Absicherung der Items auf gängige fachdidaktische Quellen im Fach zurückgegriffen. Zudem bezogen sich die Forschenden auf die Bildungsstandards (Kulturministerkonferenz [KMK], 2005), in denen Kompetenzen zum jeweiligen Fach Deutsch definiert werden (Pissarek & Schilcher, 2017). Außerdem erfolgte die Validierung über eine diskursive Aushandlung mit Expertinnen und Experten (Pissarek & Schilcher, 2015). Das Vorgehen der Validierung von Fachwissen und fachdidaktischem Wissen durch die Orientierung an Standards und/oder Normen von Expertinnen und Experten wurde auch in zahlreichen anderen Studien bestritten (u.a. König & Seifert, 2012; Rutsch, 2016; Rutsch et al., 2018a; Schumacher, 2016).

Zur Operationalisierung von fachdidaktischem Wissen: Annäherung und Rückbezug auf andere Domänen

Der Blick auf die Domäne des Orthographieerwerbs ist auch bezüglich der Operationalisierung fachdidaktischen Wissens aufschlussreich. In der erwähnten Studie von Corvacho del Toro (2013) bzw. Corvacho del Toro und Thomé (2013) wurde das fachdidaktische Wissen über vier Kompetenzfacetten erfasst, wobei auch hier auf die Facetten der COACTIV-Studie Bezug genommen wurde. Corvacho del Toro (2013, S. 130) nahm folgende Operationalisierungen vor: (1) *Wissen über die formale Beschreibung eines Rechtschreibfehlers und Lokalisierung des Fehlers*; (2) *Wissen über mögliche Erklärungen der Fehlerursache Fehlkonzeptionen, typische Fehler*; (3) *Wissen über den Lernprozess fördernde Rückmeldungen*; (4) *Wissen über geeignete Strategien bzw. Übungen zur Überwindung von Fehlschreibungen*. Diese Facetten wurden insgesamt dem diagnostischen und förderdiagnostischen Wissen zugeordnet, da die Analyse von Rechtschreibschwierigkeiten Ausgangslage für die passenden Förderangebote sei (ebd.). Diese fachdidaktischen Facetten werden in der Studie ausschliesslich durch offene Items erfasst (indem meist Hilfestellungen angegeben werden mussten) und qualitativ ausgewertet (ebd.).

Riegler und Wiprächtiger-Geppert (2016) schließen die Lücke im Bereich „orthographiebezogenes fachdidaktisches Wissen“, indem sie auch für das fachdidaktische Wissen *quantifizierbare* offene und geschlossene Items bilden. In Anlehnung an COACTIV formulieren sie jeweils orthographiebezogene Aspekte zu den drei Facetten: (1) *Erklären, Repräsentieren* [Rechtschreibstrategien, Lernhilfen zu Fehlschreibungen]; (2) *Typische Schülerfehler und Schwierigkeiten* [Fehler erkennen, Erklärungen zu Fehlschreibungen, Stand der Rechtschreibentwicklung einschätzen] und (3) *Gute Aufgaben im Rechtschreibunterricht* [Passung von Aufgaben für Lernziele, Analyse von Aufgaben und Lernmaterialien] (ebd.). Als Beispiel zur Facette „Erklären, Repräsentieren“ zeigen die Autorinnen eine Aufgabe, bei der Lehrpersonen ein Tafelbild zur Schreibung *Räuber* und *Läuferin* erstellen sollen, um dabei die morphologisch motivierte Umlautschreibung erkennbar zu machen.

Schließlich wird in der NoviS-Studie (*Novizinnen und Novizen im Schreibunterricht*) von Sturm et al. (2016), die Frage nach der Übertragbarkeit der COACTIV-Taxonomie des fachdidaktischen Wissens auf die Domäne Schreiben diskutiert. Bei NoviS wurde das fachdidaktische Wissen von Lehrpersonen zu Textprodukten von Schülerinnen und Schülern (einer Form von Fallvignette) erfasst, in dem zwei Aufgaben gestellt wurden: Zum einen musste ein Schülerinnen- bzw. Schülertext in einer offenen Frage beurteilt und das Urteil begründet werden, zum anderen mussten diese Texte kommentiert und eine Rückmeldung im Sinne von Verbesserungstipps gegeben werden. Diese textbezogenen Urteile und Kommentare wurden

anschließend in quantifizierbare Codes umgewandelt (ebd.). In der Studie wird das Beurteilen von Textprodukten und die formative Rückmeldung zu Leistungen schwacher Schreiberinnen und Schreiber dem fachdidaktischen Wissen zugeordnet und dieses Wissen wird – wie bereits bei Corvacho del Toro – in die Nähe diagnostischer Kompetenzen gerückt (Sturm, 2016). Eine Analogie zum vorliegenden Projekt besteht darin, dass das diagnostische Wissen u.a. mittels audiobasierter Fallvignetten zu Leseschwierigkeiten erfasst wurde.

Bezüge zum fachdidaktischen Wissen im Bereich Lesen aus dem deutschsprachigen und dem angloamerikanischen Raum

Anregungen zur Operationalisierung fachdidaktischen Wissens in der Domäne Lesen finden sich in der Studie von Rutsch (2016). Sie operationalisiert das lesedidaktische Wissen von Lehramtsstudierenden der Sekundarstufe I in die zwei Facetten „Förderwissen/Intervention“ und „Diagnose/Inhaltswissen“ (ebd., S. 84f.). Die Facette *Förderwissen/Intervention* wird auf den Aspekt „Wissen über Erklären und Darstellen“ von Shulman (1986; 1987) bezogen und umfasst die Fähigkeit, adäquat auf Verständnisschwierigkeiten beim Lesen eines Textes zu reagieren und Hilfestellungen zu geben. Inhaltlich wird hier Bezug genommen auf die aktuelle Theorie und Empirie im Bereich Leseflüssigkeit, Lesemotivation und Lesestrategien. Die Facette *Diagnose/Inhaltswissen* wird in Verbindung mit Shulmans (1986; 1987) Facette „Wissen über fachbezogene Schülerkognitionen“ gebracht und umfasst das Wissen um die zuverlässige Diagnose von Leseleistungen der Schülerinnen und Schüler. Dafür werden insbesondere Kenntnisse über mentale Abläufe während des Leseprozesses sowie das Wissen über Einschränkungen in Teilbereichen des Lesens benötigt. Daneben wird das Wissen über schwierigkeitsgenerierende Textmerkmale erfragt (Rutsch, 2016). Erfasst wird dies mittels schriftliche Fall-Vignetten über unterrichtsnahe lesedidaktische Probleme, die durch die Studienteilnehmenden in einer sechsstufigen Likert-Skala beurteilt werden müssen. Die beiden Facetten korrelieren stark miteinander, so dass hier aufgrund der empirischen Daten letztlich eher von *einer* Dimension ausgegangen wurde (ebd.).

Ein weiterer Bezug zur Domäne Lesen auf Primarstufe findet sich in der Studie aus den USA von Phelps und Schilling (2004). Zum fachdidaktischen Wissen wurden über geschlossene Antwortformate die zwei Facetten „Knowledge of students and content“ (KSC) und „Knowledge of teaching and content“ (KTC) gebildet. Diese wurden jeweils den Bereichen „Comprehension“ und „Word Analysis“ zugeordnet (ebd.). Operationalisiert wurden die Facetten folgendermaßen:

1) *Knowledge of students and content (KSC)* betrifft das Wissen darüber, wie Arbeits-Produkte von Schülerinnen und Schülern im Bereich Lesen entschlüsselt und interpretiert werden müssen. Wenn ein Kind einen Text ungenau liest, muss dieser Fehler zunächst einmal erkannt werden. Bei dieser Facette steht das „Interpretieren“ von (Fehl-) Leistungen der Schülerinnen und Schüler im Fokus, d.h. Lehrpersonen müssen Fehler erkennen, richtig interpretieren, aber noch keine Entscheidungen über allfällige Hilfestellungen fällen (ebd.).

2) *Knowledge of teaching and content (KTC)* fokussiert auf das Lehren, bezeichnet als „Teaching students“. Gefragt wird hier u.a. danach, was Lehrpersonen einem Kind sagen, wenn es über ein schwieriges Wort stolpert, oder welche Hilfestellungen Lehrpersonen geben, um eine schwierige Textpassage zu verstehen (ebd.). Die Erfassung erfolgt über die Beschreibung einer lesebezogenen Problemstellung und einer anschließenden Auswahl richtiger und falscher Antworten im Multiple-Choice-Format.

3.3.1.3 Diagnostisches Wissen im Kontext des Professionswissens

Beschreibung und Herausforderungen zum diagnostischen Wissen

Wie aus vorangegangenen Erläuterungen hervorgeht, wird das diagnostische Wissen im Zusammenhang mit dem fachdidaktischen Wissen meist implizit (teilweise aber auch explizit) erwähnt. Vor dem Hintergrund des entwickelten Befragungsinstruments, bei dem neben Grundlagenwissen zum Leserwerb und Lesen auch das diagnostische Wissen bei Leseschwierigkeiten erfasst wird, scheint die Klärung seiner Operationalisierung und theoretischen Einordnung von Bedeutung zu sein.

Baumert und Kunter (2006, S. 489) halten fest, dass in der gegenwärtigen Diskussion über Lehrerinnen- und Lehrerbildung kaum eine Komponente professionellen Wissens und Könnens so prominent sei wie die so genannte „diagnostische Kompetenz“. Sie wird bisweilen als „Schlüsselkompetenz“ für das Unterrichten bezeichnet (Moser Opitz & Nührenbörger, 2015, S. 491).

Bemerkenswert ist, dass das diagnostische Wissen in den erwähnten Modellen nicht als eigener Bereich bzw. als Facette beschrieben wird, und es stellt sich die Fragen, wo das diagnostische Wissen in Professionsmodellen situiert werden kann, da „aus empirischer Perspektive noch nicht geklärt ist, ob diagnostische Kompetenz einen Aspekt des fachdidaktischen Wissens, eine Kombination aus fachlichem und fachdidaktischem Wissen oder vielmehr eine tatsächliche abgrenzbare Fähigkeit darstellt“ (Sturm, 2016, S. 118). Eine abschließende Antwort hierzu ist derzeit nicht möglich. Dennoch soll hier ein kurzer Blick auf den theoretischen Diskurs geworfen werden.

Diagnostische Kompetenz lässt sich gemäß Weinert (2000) wie folgt umschreiben:

Dabei handelt es sich um ein Bündel von Fähigkeiten, um den Kenntnisstand, die Lernfortschritte und die Leistungsprobleme der einzelnen Schüler sowie die Schwierigkeiten verschiedener Lernaufgaben im Unterricht fortlaufend beurteilen zu können, so dass das didaktische Handeln auf diagnostische Einsichten aufgebaut werden kann. (S.14f.)

Mit Bezug auf die genannte Definition verwendet Helmke (2017, S. 119) ausdrücklich den Begriff „diagnostische Expertise“ und grenzt ihn vom Begriff „diagnostische Kompetenz“ ab, da mit Letzterem meist nur auf die Urteilsgenauigkeit (*accuracy*) Bezug genommen werde. Expertise ist umfassender und beinhaltet auch *methodisches* und *prozedurales* Wissen (Wissen über Methoden zur Einschätzung von Schülerinnen- und Schülerleistungen) sowie *konzeptionelles* Wissen (Urteilstendenzen, Urteilsfehler) (ebd.). Eine weitere Beschreibung findet sich bei von Aufschnaiter et al. (2015), die zur diagnostischen Kompetenz – im Sinne von Anhaltspunkten für die Professionalisierung von Lehrpersonen – folgende Maßnahmen und Aktivitäten zählen:

- 1) Kenntnis über grundlegende diagnostische Verfahren und Tests [...]
- 2) Entwicklung eigener Tests und Lernstandserhebungen, die Aufschluss über Schwächen, Misskonzepte und Stärken der Lernenden geben und ggf. Unterschiede zwischen Schülern in der Bearbeitung verdeutlichen
- 3) Analyse der Qualität von selbst entwickelten Tests
- 4) Interpretation von Schülerdaten aus Leistungstests und Lernstandserhebungen
- 5) Entwicklung und Erprobung diagnostischer Lehrerfragen
- 6) Analyse von Schülerdokumenten. (S. 743)

Das Ziel von Diagnoseleistungen in diesem Verständnis ist es, Informationen über Lernergebnisse, Lernvoraussetzungen und Lernvorgängen der Schülerinnen und Schüler zu gewinnen und für pädagogische Entscheidungen verschiedener Art zu nutzen (Schrader, 2013). Damit wird deutlich, dass Diagnostik nicht dem Selbstzweck dient, sondern dass sich aus der Feststellung von Merkmalen und deren Ausprägung Aussagen über anschließende adaptive Maßnahmen ableiten lassen sollten (von Aufschnaiter et al., 2015).

Nebst den erwähnten aufgabenbezogenen Definitionen wird diagnostische Kompetenz in der empirischen Forschung oft enger gefasst im Sinne von „Urteilsakkuratheit“. Die Definition der Urteilsakkuratheit beschäftigt sich mit der Frage, wie gut Lehrpersonenurteile mit den bei Schülerinnen und Schülern gemessenen „tatsächlichen“ Merkmalsausprägungen übereinstimmen (Schrader, 2013, S. 157). Dabei wird die Rangordnungskomponente oft als das eigentliche Maß für die diagnostische Kompetenz (Sensitivität) angesehen, da die so gemessene diagnostische

Kompetenz im Gegensatz zu anderen Komponenten (Niveau- und Differenzierungskomponente) mit der Lernentwicklung korreliert (Karst, 2017; Schrader, 2013). Je höher die Schülerinnen- und Schülerleistung mit dem Urteil der Lehrperson korreliert, umso besser kann diese die Rangreihe ihrer Schülerinnen und Schüler einschätzen, d.h. umso besser ihre diagnostische Kompetenz (Karst, 2017). Bei dieser Definition werden allerdings die diagnostischen Anforderungen in unterrichtlichen Situationen außer Acht gelassen. Es wird von einer fach- und kontextunspezifischen Kompetenz ausgegangen, die sich auf die „Genauigkeit“ als Gegenstand diagnostischer Kompetenz bezieht (von Aufschnaiter et al., 2015, S. 740). Die Fähigkeit genaue Urteile abzugeben, scheint aber nur einen Teilaspekt der diagnostischen Kompetenz abzubilden (Schrader, 2014).

Damit ist die Frage angesprochen, ob diagnostisches Wissen domänenspezifisch und damit (fach-)didaktisch verankert ist, oder ob es als fächerübergreifende Kompetenz zu betrachten sei (Baumert & Kunter, 2006; Moser Opitz & Nührenbörger, 2015). Letzteres wird allerdings eher bezweifelt, da bisher nicht bekannt ist, ob Lehrpersonen, die verschiedene Fächer unterrichten, über die verschiedenen Fächer hinweg konsistent urteilen oder nicht (Baumert & Kunter, 2006). „Plausibilitätsargumente“ sprechen dafür, dass Leistungsdiagnostik in gewissen Fächern vom fachdidaktischen Wissen der Lehrpersonen abhängen könnte (ebd., S. 489). Die Sichtweise, dass eher nicht von einem globalen Konstrukt diagnostischer Kompetenz auszugehen ist und sich Diagnosekompetenz je nach Unterrichtsfach und einzuschätzendem Merkmal unterscheidet, wird auch empirisch gestützt (Karst, 2017).

Wird die Diagnosekompetenz nicht auf die „Urteilsgenauigkeit“ reduziert, so lässt sich ableiten, dass vom theoretischen Standpunkt aus eine fach-, d.h. domänenspezifische Konzeptualisierung diagnostischer Kompetenz plausibel erscheint (Artelt, 2009). Exemplarisch zeigt sich dies im Fach Mathematik, wo deutlich wird, dass Diagnosekompetenz auf verschiedenen Ebenen ansetzt und Wissen über passende Aufgabenstellungen, die Analyse von Fehlermerkmalen sowie die Kenntnis mathematischer Entwicklungsprozesse notwendig ist (Moser Opitz & Nührenbörger, 2015). Übertragen auf den Fachbereich Lesen zeigt sich die domänenspezifische diagnostische Kompetenz darin, dass Lehrpersonen Aufgabenschwierigkeiten einschätzen müssen, Kenntnisse über lesebezogene Diagnoseverfahren besitzen sollten, Vorstellungen über Fähigkeiten der Schülerinnen und Schüler im Bereich Lesen haben und im nächsten Schritt geeignete Möglichkeiten zur Leseförderung auswählen können (Schmidt, 2015). Zur lesebezogenen diagnostischen Kompetenz gehört zudem die Fähigkeit, sowohl gelingende als auch misslingende Lese- und Textverstehensprozesse zu diagnostizieren. Auch müssen Lehr-

personen über schwierigkeitsgenerierende Textmerkmale Bescheid wissen (Artelt, 2009; Bertschi-Kaufmann, 2011; Rutsch, 2016).

Verortung diagnostischer Kompetenz im Rahmen von Professionsmodellen

Es stellt sich die Frage, wie die diagnostische Kompetenz in Professionsmodellen zu verorten ist, d.h. ob das Konstrukt als eigenständiges Konstrukt konzeptualisiert ist oder ob es bei der Erforschung professioneller Kompetenz „mitläuft“ (von Aufschnaiter et al., 2015, S. 742). Von Aufschnaiter et al. (2015) stellen fest, dass die Mehrzahl der Ansätze diagnostische Kompetenz in einzelnen Dimensionen der Modellierung professioneller Kompetenz verortet bzw. dass sie *quer* zu den Dimensionen liegt. „Diese Überlegungen tragen der Annahme Rechnung, dass diagnostische Kompetenz zwar eine Kernkompetenz von Lehrkräften ist, darüber hinaus jedoch eine Reihe weiterer Kompetenzen die Professionalität von Lehrkräften ausmachen und deshalb eine entsprechend gebündelte Beschreibung erforderlich ist“ (ebd., S. 742).

Schrader (2014) verortet die diagnostische Kompetenz im Bereich pädagogischen *oder* fachdidaktischen Wissens und verweist hier auf die Typologisierung von Shulman (1987) bzw. das daraus abgeleitete Modell von Baumert und Kunter (2006). Auch an anderen Stellen wird die Nähe des Konstrukts *diagnostisches Wissen* zum Wissensbereich *fachdidaktisches Wissen* angenommen (Corvacho del Toro, 2013; Rutsch, 2016; Sturm, 2016). Im COACTIV-Modell ist dieser Bezug in den Beschreibungen der einzelnen fachdidaktischen Facetten „Diagnostik von Schülerinnen- und Schülerwissen“ und „Wissen über das diagnostische Potenzial von Aufgaben“ (Baumert & Kunter, 2011a, S. 37f.) expliziert. Baumert und Kunter (2006) vermuten, dass das *fachdidaktische Können* gerade für die Auswahl von Arbeitsaufträgen und Aufgaben eine besondere Bedeutung hat, da gut gewählte Aufgaben ein großes diagnostisches Potenzial in sich selbst tragen. Neben der Nähe zum fachdidaktischen Wissen wird aber auch ein Bezug zum *Fachwissen* hergestellt, da diagnostische Urteile u.a. auf fachlichem Wissen aufbauen und Aussagen dazu, was Aufgaben schwer macht bzw. welche Strategien zur Bewältigung notwendig sind, fundiertes Wissen über Lösungsmöglichkeiten und Anforderungen voraussetzt (Artelt, 2009). Diagnostische Kompetenz „lässt sich als Teil des fachdidaktischen Wissens darstellen und stellt somit nach Shulman (1987) neben dem Fachwissen und dem pädagogischen Wissen das dritte Standbein professioneller Handlungskompetenz von Lehrerinnen und Lehrern dar“ (Artelt, 2009, S. 125 f.).

Neben dieser theoretischen Verortung diagnostischer Kompetenz lohnt sich der Blick auf die Empirie. Im Zuge der empirischen Prüfung des Modells der Michigan-Gruppe (Hill et al., 2004) hat sich auf Grundlage der exploratorischen Faktorenanalyse gezeigt, dass neben zwei

inhaltlichen Dimensionen ein Faktor gebildet werden konnte, der vor allem schülerinnen- und schülerbezogene diagnostische Kompetenzen umfasst (Baumert et al., 2010). Dieser Befund lässt den Schluss nahe, diagnostische Kompetenz auch als eigenständige Dimension zu fassen.

3.3.2 Zum Verhältnis von Fachwissen und fachdidaktischem Wissen/Können

Fachwissen und fachdidaktisches Wissen stehen in einem direkten Zusammenhang, wobei sich das Verhältnis zum einen theoretisch-konzeptionell, zum anderen aber auch messmethodisch und empirisch beurteilen lässt.

Zusammenhang und Unterscheidung von Fachwissen und fachdidaktischem Wissen/Können

Theoretisch-konzeptionell berufen sich die meisten Studien auf die Wissenstaxonomie von Shulman, der Fachwissen und fachdidaktisches Wissen als unterscheidbare, aber sich überlappende Wissensbereiche beschreibt (Pissarek & Schilcher, 2017). Die Frage nach der Trennbarkeit wurde in zahlreichen Studien und bezogen auf unterschiedliche Fächer in den bereits mehrfach zitierten Studien COACTIV (Kunter et al., 2011a), TEDS-LT (Blömeke et al., 2013a), FALKO (Krauss et al., 2017a) untersucht. Einerseits wird von einer Trennbarkeit ausgegangen, aber auch ein theoretischer Zusammenhang postuliert und fachdidaktisches Wissen entsprechend als eine *spezifische* Form mathematischen Wissens bezeichnet (Baumert & Kunter, 2011b, S. 185). Das Verhältnis stellt sich so dar, dass fachdidaktisches Wissen auf dem Fachwissen basiert, aber durch dieses nicht substituiert werden kann. Fachdidaktisches Wissen ist im Unterschied zum Fachwissen in der Qualität des Unterrichtsprozesses nachweisbar (ebd.). Während in oben genannten Studien Fachwissen und fachdidaktisches Wissen bereits bei der theoretischen Konzeptualisierung unterschieden wurden, wurde etwa im Modell der Michigan-Gruppe (Hill et al., 2004) eine engere Verbindung angenommen (Baumert & Kunter, 2011b). Auch in einem alternativen Modell von Lindmeier (2011) wird eine getrennte Analyse der beiden Bereiche zugunsten einer „Basic Knowledge Component“ schon in der Konzeptualisierung fallengelassen und mit den Worten kommentiert: „Why (artificially) distinguish between constituents of a knowledge area that are strongly interrelated and only to be separated with difficulty?“ (S. 105).

Die hier angesprochene enge Verzahnung zeigt sich auch im Zusammenhang mit messmethodischen Überlegungen. Neuweg meint, dass die Frage der Trennbarkeit von Fachwissen und fachdidaktischem Wissen keine rein empirische Frage sei, sondern wesentlich von der Frage der Ansprüche an das Fachwissen selbst und damit auch von der Art ihrer Messung abhängen (Neuweg, 2014). Neuweg (ebd., S. 591) spricht von der impliziten fachdidaktischen Dimension des Fachwissens oder auch von einer „Brüchigkeit“ der Unterscheidung der beiden

Dimensionen, da die Versuchsperson in jedem Fachwissenstest dazu genötigt werde, das eigene Wissen zu repräsentieren, und da ein Test zudem vor dem Hintergrund konkreter Vorstellungen entwickelt werde, was als gehaltvolle Repräsentation von Fachwissen gelten könne. Sobald das eigene Wissen – als ein wichtiger Bestandteil von fachdidaktischem Wissen – dargestellt werden soll, lässt es sich sowohl theoretisch als auch praktisch nur noch schwer vom „reinen“ Fachwissen unterscheiden (ebd.). Trotz dieser – auch methodisch bedingten – starken Überschneidung der Bereiche Fachwissen und fachdidaktisches Wissen gilt die Trennung der Komponenten dennoch gemeinhin als akzeptiert. Die getrennte Skalierung von Fachwissen und fachdidaktischem Wissen findet ihre Begründung wohl auch darin, dass die Wissensvermittlung und der Erwerb fachdidaktischen Wissens in zahlreichen Ländern (auch in Deutschland) weitgehend unabhängig voneinander erfolgen (Stancel-Piatak et al., 2013). Die getrennte Operationalisierung dürfte damit auch mit den Ausbildungsstrukturen zusammenhängen. Anders stellt sich die Situation in der Schweiz dar, da Primarlehrpersonen ihr Wissen im Rahmen der Lehrerinnen- und Lehrausbildung meist an pädagogischen Hochschulen erwerben, wobei die Fachwissenschaft und die Fachdidaktik meist durch dieselbe Person vermittelt wird (Oser, 2015). In welcher Weise sich die Abhängigkeit vor dem Ausbildungshintergrund allerdings gestaltet, kann aufgrund des uneinheitlichen Forschungsstands nicht abschließend beantwortet werden (Jandl, 2016).

Empirische Befunde

Mit Blick auf die Empirie werden vor allem die Daten der COACTIV-Studie wiederkehrend referiert. Für den Bereich der Mathematik wird bei faktoranalytischer Trennbarkeit der Bereiche fachliches und fachdidaktisches Wissen eine Korrelation von $r = .79$ berichtet. Somit kann davon ausgegangen werden, dass sich die beiden Bereiche zwar überlappen, aber dass dennoch von zwei unterschiedlichen Wissensbereichen ausgegangen werden kann (Krauss et al., 2011, S. 148). Helmke (2012, S. 112) kommentiert den Zusammenhang dahingehend, dass die Bereiche eng zusammenhängen, die Leistung der Schülerinnen und Schüler (hier im Fach Mathematik) aber ausschließlich auf das *fachdidaktische Wissen* zurückgehe. Dabei ist das Fachwissen aber nicht unbedeutend, da Mängel im Fachwissen die Entwicklung fachdidaktischer Ressourcen limitieren (Baumert & Kunter, 2011b). Lücken im Fachwissen können durch das fachdidaktische Wissen kaum kompensiert werden (ebd.). Zu den wichtigsten Ergebnissen aus qualitativen Studien gehört zudem der interpretative Nachweis, dass das verfügbare fachdidaktische Handlungsrepertoire von Lehrpersonen weitgehend von der Breite und Tiefe ihres konzeptionellen Fachverständnisses abhängt (Baumert & Kunter, 2006). Dies bedeutet, dass das

fachdidaktische Wissen nicht einfach im Fachwissen aufzugehen scheint, sondern eine eigene Wissenskomponente „sui generis“ zu sein scheint (ebd., S. 493).

Die Frage nach der empirischen Abgrenzbarkeit der Wissensbereiche scheint auch von den untersuchten Fächern abhängig zu sein. So zeigt sich, dass in den Fächern Deutsch und Englisch nur mittlere latente Zusammenhänge zwischen Fachwissen und fachdidaktischem Wissen gemessen wurden, wenn das fachdidaktische Wissen weniger stoffnah und weniger anforderungsnah, dafür aber ausbildungsorientierter (Niveau „Erinnern“) erfasst wurde (Neuweg, 2014). Belege für diese These finden sich im Vergleich der Daten im Fach Deutsch von TEDS-LT, $r = .49$ (Bremerich-Vos et al., 2011, S. 62) und FALKO-Deutsch, $r = .74$ (Pissarek & Schilcher, 2017, S. 99). In TEDS-LT wurde vorwiegend deklaratives ausbildungsbezogenes Wissen über mehrheitlich geschlossene Items erfasst (Pissarek & Schilcher, 2017). Dagegen war das Ziel der FALKO-Tests, die unterrichtliche Handlungsfähigkeit abzubilden, was zu einer starken Bevorzugung offener Items (insbesondere beim fachdidaktischen Wissen) geführt hat (Krauss et al., 2017b). In allen mit FALKO getesteten Fächern zeigt sich das Bild von trennbaren, aber sich überlappenden Bereichen, wobei die Korrelationen zwischen $r = .56$ (Englisch) und $r = .86$ (Physik) liegen.

Mit Bezug auf Tests aus der Domäne Schriftspracherwerb zeigt sich zum einen, dass in der Studie von Phelps und Schilling (2004) von trennbaren Dimensionen „Fachwissen“ (*Knowledge of Content*) und „fachdidaktische Wissen“ (*Knowledge of Teaching and Content*) auszugehen ist. Schließlich zeigt der Blick auf die Studie aus dem Bereich des Orthographieerwerbs von Riegler und Wiprächtiger-Geppert (2016), dass der Versuch, aus fachwissenschaftlicher und fachdidaktischer Sicht plausible Aufgaben für entsprechende Skalen zu bilden, vorerst gescheitert ist. Die Autorinnen führen dies u.a. auf die Heterogenität der Aufgabentypen zurück. Insgesamt ist die Forschungslage zur Dimensionierung von Wissensbeständen im Bereich Lesen äußerst dünn. Unter Rückbezug auf andere Fächer und Domänen zeigt sich zudem, dass sich das Testformat, die Domäne und die Konzeptualisierung (akademisches vs. praktisches Wissen) auf das Verhältnis von Fachwissen und fachdidaktischem Wissen auswirken.

3.3.3 Mentale Repräsentationen von Professionswissen und messmethodische Folgerungen

Wird Kompetenz im Sinne einer „Verbindung von Wissen und Können in der Bewältigung von Handlungsanforderungen“ (Klieme & Hartig, 2008, S. 19) verstanden, so stellt sich die Frage, wie Wissen und Können zu beschreiben sind. Die Auseinandersetzung mit der Beschaffenheit verschiedener Wissensformen bis hin zur Handlung im Kontext der Professionsforschung ist Grundlage, um Wissen und Können später in geeigneter Form erfassen zu können. Im

Folgenden werden die Begriffe *Wissen* und *Können* sowie der Begriff *Handlung* in Beziehung gesetzt und mit verwandten Konzepten abgeglichen. Da sich Wissen und Können sowie verwandte Umschreibungen in verschiedenen Taxonomien fassen lassen, ist eine einheitliche Beschreibung äußerst schwierig. Zum einen liegen nach Vogelsang und Reinhold (2013) Konzepte von *Transfervorstellungen* vor, „nach denen systematisch erworbenes, explizierbares Wissen in Können transferiert wird“ (S. 322), und zum anderen gibt es Konzepte, die eher *Differenzvorstellungen* entsprechen (ebd.). Demzufolge zeigen sich deutliche Unterschiede zwischen dem systematischen Ausbildungswissen und dem unterrichtlichen Handlungswissen. Alle bestehenden theoretischen Konzepte von Wissen und Können lassen sich auf einem gedachten Kontinuum zwischen diesen beiden Polen positionieren (ebd.). Im Folgenden werden Transfervorstellungen wie sie in der Expertiseforschung angenommen werden, in den Fokus genommen, und es wird versucht, eine Annäherung an die Konstrukte Wissen und Können bzw. Formen mentaler Repräsentationen von Wissen zu finden.

3.3.3.1 Differenzierung von Wissen und Können

Das Wissen von Lehrpersonen ist gemäß Neuweg (2014) zum Fokusbegriff geworden, „in dem das Interesse am Lernen, am (expliziten) Wissen *und* Können von Lehrern zueinanderfinden“ (S. 583). Damit wird klar, dass Wissen und Können in einem *Verhältnis* zueinander stehen oder – anders ausgedrückt – dass sich die Konstrukte graduell voneinander unterscheiden. Bei Baumert & Kunter (2006) werden die Begriffe Wissen und Können mit den Begriffen deklaratives und prozedurales Wissen verknüpft: „Es besteht weitgehende Übereinstimmung darüber, dass Wissen und Können – also deklaratives, prozedurales und strategisches Wissen – zentrale Komponenten der professionellen Handlungskompetenz von Lehrkräften darstellen“ (ebd., S. 481). In Anlehnung an König (2010, S. 61) und Bromme (1992, S. 121) könnte man beim deklarativen Wissen von „wissen, dass“ und beim prozeduralen Wissen vom „wissen, wie“ sprechen. Auch wenn diese einfache Unterscheidung bisweilen als unpräzise kritisiert werden kann, hat sie sich als pragmatische Einteilung in der Diskussion um eine Kategorisierung mentaler Repräsentation von Professionswissen dennoch durchgesetzt (König, 2010).

1) *Deklaratives Wissen* umfasst die „Gesamtheit der Kenntnisse einer Person über Fakten, Sachverhalte, Personen u.a. (...) es gilt als verbalisierbar und nicht situationsspezifisch; repräsentiert ist es in propositionaler Form“ (König, 2010, S. 62). Diese Art von Wissen umfasst bewusste Denkinhalte und hat einen vergleichsweise geringen Handlungsbezug. Es stellt aber dennoch eine Voraussetzung für prozedurales Wissen dar (Brühwiler et al., 2017).

2) *Prozedurales Wissen* beinhaltet vor allem vernetztes Wissen im Sinne von gespeicherten Schemata, semantischen Netzwerken oder Hierarchien (König, 2010). Indem das deklarative Wissen prozeduralisiert wird, erreicht es eine höhere Kompetenzstufe und wird dabei in Können überführt (Klieme et al., 2007). Da es meist auch unbewusste kognitive Operationen beinhaltet (Brühwiler et al., 2017), können prozedurale Wissensselemente allerdings nur in dem Masse berücksichtigt werden, wie es das jeweilige Messverfahren auch erlaubt (König, 2010). Implizites, handlungsrelevantes Wissen muss in entsprechend geeigneter Form hervorgerufen werden, etwa durch die Beurteilung von Textvignetten (Brühwiler et al., 2017).

Dem Begriffspaar verwandt sind nach König (2010) auch die Begriffe des theoretisch-formalen Wissens (deklarativ) und des praktischen Wissens (prozedural). Die Unterscheidung zwischen *formal knowledge* und *practical knowledge* wird auch im Zusammenhang aktueller Modellierungen von Professionswissen bzw. zur Abgrenzung von Wissensbereichen verwendet (u.a. Krauss et al., 2017b; Rutsch et al., 2018c). Das theoretisch-formale Wissen umfasst vorrangig das fachliche Wissen, enthält aber auch Teile des fachdidaktischen und psychologisch-pädagogischen Wissens, während unter praktischem Wissen und Können das erfahrungsbasierte Wissen von Lehrpersonen verstanden wird, das auf konkrete Problemstellungen im Unterricht bezogen und damit kontextspezifisch ist (Rutsch et al., 2018c).

Messmethodische Folgerungen

Unter messmethodischem Blickwinkel hat diese Kategorisierung zur Folge, dass mit Papier-Bleistift-Verfahren vor allem theoretisch-formales, deklaratives Wissen sowie explizierbares praktisches, prozedurales Wissen erfasst werden können (König, 2010). Das Handeln im Klassenzimmer entzieht sich einer Papier-Bleistift-Testung weitgehend, so dass hier auf andere methodische Zugänge zurückgegriffen werden muss (ebd.). Während also das deklarative, explizite Wissen oft über Multiple-Choice-Fragen erfasst wird, geschieht dies beim prozeduralen Wissen meist über Operationen und offene Aufgaben, bei denen die Teilnehmenden selber etwas tun müssen (Corvacho del Toro & Thomé, 2013). Auch in den FALKO-Tests wurde durch die Anwendung offener Itemformate – vor allem beim fachdidaktischen Wissen – versucht, unterrichtliche Handlungsfähigkeit abzubilden (Krauss et al., 2017b). Hier wurde über unterrichtsnahe typische Aufgaben nach Erklärungen zu Fachinhalten (*Welche Hilfestellungen können in einem konkreten Fall helfen, einen Fehler zu überwinden*) und Interpretationen von Schülerkognitionen (*Welche Hauptprobleme liegen bei einem Kind bei einer falsch gelösten Aufgabenstellung vor*) gefragt (Pissarek & Schilcher, 2017, S. 77ff.). Eine weitere Annäherung an eine kontextualisierte und situationsspezifische Erfassung professionellen Wissens und

Könnens kann zudem über schriftliche oder videographierte Unterrichtsvignetten erfolgen, wie beispielsweise im Rahmen der EKoL-Studie „Effektive Kompetenzdiagnose im Unterricht“ (Rutsch et al., 2018c).

3.3.3.2 Kognitive Anforderungen von Wissen und Können

Neben den bereits unterschiedenen Begriffen ist eine weitere Kategorisierung von Wissensbereichen im Sinne kognitiver Anforderungsniveaus hilfreich, da sie sich dem Phänomen des Wissens und Könnens explizit aus der Perspektive messtheoretischer Anforderungen annähert. König (2012, S. 145f.) stellt hierzu eine dreiteilige Taxonomie vor. Es sind dies die drei Dimensionen *Erinnern*, *Verstehen* und *Kreieren*, die je eigene Qualitäten von Wissen beinhalten. Die drei Anforderungsniveaus werden neben der Inhaltsdimension oft für die Beschreibung der kognitionsbezogenen Struktur von Testinstrumenten verwendet, wobei Brühwiler et al. (2017, S. 213) anstelle des Kreierens den Begriff *Anwenden* brauchen.

Die drei genannten kognitiven Dimensionen wurden in der TEDS-M-Studie (Blömeke et al., 2010a) sowie der TEDS-LT-Studie (Blömeke, 2011; Bremerich-Vos et al., 2011) der Entwicklung der Testitems zugrunde gelegt. In der TEDS-LT-Studie wurden verschiedene kognitive Anforderungsniveaus beschrieben, die aus pragmatischen Gründen in drei Gruppen zusammengefasst wurden: *Erinnern/Abrufen*; *Verstehen/Anwenden*; *Bewerten/Kreieren* (Bremerich-Vos et al., 2011). Für die Konstruktion der Testaufgaben ist diese Unterscheidung wichtig, weil sie Aussagen darüber zulässt, auf welcher Ebene das zur Verfügung stehende Wissen der befragten Personen anzusiedeln ist.

1) *Erinnern/Abrufen*: Diese Dimension des Wissens bezieht sich in erster Linie auf deklaratives Wissen, das in annähernd identischer Form auch gelernt und wiedergeben werden kann (Blömeke, 2011). Beim Erinnern/Abrufen wird Wissen benötigt, das zum Lehrbuchwissen oder Erfahrungswissen zählt und bei dem es um das korrekte Erkennen und Benennen von Sachverhalten geht (König, 2012). Testaufgaben dieses Typs erfordern von der befragten Person, dass sie Beispiele für Definitionen geben kann, dass sie fähig ist, zu einem Begriff, Phänomen oder Konzept Elemente aufzählen und dass sie Begriffe und Konzepte erkennt (ebd.). In diesem Zusammenhang wird auch der Begriff des *Kennens* eingeführt (Döhrmann, Kaiser & Blömeke, 2010, S. 172), und es zeigen sich Überschneidungen von theoretisch-formalem Wissen und deklarativem Wissen. Der kognitive Aspekt des „Reproduzierens“ wird zudem meist dem Fachwissen zugeordnet, womit es hier zur Überlagerung von kognitiven Prozessen mit einer Dimension des Professionswissens kommt (vgl. Modell zur Analyse und Konzeption von Aufgaben in der Lehrerbildung Reintjes, Keller, Jünger & Düggele, 2016).

2) *Verstehen/Analysieren*: Neben dem Erinnern von Wissen geht es auf dieser Stufe zudem um die Verbindung dieses Wissens mit einer Problemstellung. In der Studie TEDS-LT wird deshalb der Begriff des *Anwendens* dem Verstehen zugeordnet, wobei Anwenden den Gebrauch von Prozeduren zum Üben oder zur Problemlösung meint (Bremerich-Vos et al., 2011). Testaufgaben, die diese Qualität von Wissen erfassen, erfordern von den Testpersonen, dass sie Konzepte und Sachverhalte charakterisieren oder erklären können (König, 2012). Weiter müssen sie auf dieser Stufe Begriffe und Konzepte auswählen, um damit eine Erklärung zu liefern. Es geht aber auch darum, einen Sachverhalt, einen Fall oder Begriffe zu vergleichen, zu kategorisieren, zu ordnen und zu interpretieren (ebd.). Hier werden also Aspekte prozeduralen Wissens bzw. Könnens angesprochen.

3) *Bewerten/Kreieren*: Auf dieser Qualitätsstufe kommt hinzu, dass das Wissen auch genutzt werden kann, um Handlungsoptionen zu entwickeln und zu formulieren (König, 2012). Deshalb wird auch von „Generieren von Handlungsoptionen“ gesprochen (Blömeke, 2011, S. 16). Handlungsoptionen sind relevant für die Bewältigung einer typischen Situation. Dazu kommt, dass das praktische Wissen und Können als mögliche mentale Repräsentation beschrieben und dargelegt werden kann (König, 2012). Mit dem Begriff des Bewertens ist eine Beurteilung im Hinblick auf die Qualität sowie die Richtigkeit und Angemessenheit der Problemlösung gemeint (Bremerich-Vos et al., 2011). Im erwähnten Modelle von Reintjes et al. (2016) kommt es auch in diesem Bereich zu einer Überlagerung der kognitiven Prozesse und der Dimensionen des Professionswissens. *Analysieren/Entwickeln/Transformieren/ Bewerten* werden hier in die Nähe der Dimension „fachdidaktisches Wissen“ gerückt.

Die erwähnten kognitiven Prozesse sind nicht zwingend hierarchischer Natur (Bremerich-Vos & Dämmer, 2013), da dies erfordern würde, dass das zugrunde liegende Wissen (inhaltlich) exakt denselben Schwierigkeitsgrad aufweist (Blömeke, 2011). Es kann also nicht davon ausgegangen werden, dass Items auf der Ebene des Erinnerns und Abrufens einfacher sind als Items auf der Ebene des Verstehens oder Kreierens (Bremerich-Vos et al., 2011). So können komplexe Erinnerungsaufgaben durchaus schwieriger sein als leichte Aufgaben zum Kreieren (Blömeke, 2011). In diesem Sinne müssen die Dimensionen eher im Sinne verschiedener Subdimensionen gedacht werden und nicht als Hierarchie kognitiver Prozesse (Bremerich-Vos et al., 2011).

3.4 Wirkung des Professionswissens im Unterrichtskontext

Die Entwicklung eines Befragungsinstrumentes zur Erfassung von Fachwissen und fachdidaktischem Wissen macht vor allem dann Sinn, wenn Hinweise vorliegen, die einen Zusammenhang von gemessenem Professionswissen und Schülerinnen- und Schülerleistungen vermuten lassen. Im Folgenden soll ein Blick auf diesen Zusammenhang geworfen und damit aufgezeigt werden, über welche Wege Professionswissen letztlich *lernwirksam* wird und wie dieses Wissen zu erfassen ist (Brühwiler et al., 2017). Um diesen Zusammenhang in einen Kontext zu stellen, soll zunächst die Wirkungskette schulischer Bildung in den Blick genommen werden. Dafür werden zwei Schnittstellen genauer analysiert. Erstens wird gezeigt, wie sich das professionelle Wissen von Lehrpersonen im Rahmen der Ausbildung und der späteren Berufspraxis entwickelt, und zweitens werden Zusammenhänge dieses Wissens im Hinblick auf die Lernerfolge der Schülerinnen und Schüler betrachtet. Da diese Schnittstellen im Bereich Lesen bis anhin nur ungenügend untersucht worden sind, werden hier Studien aus anderen Fächern und Domänen mitberücksichtigt, im Wissen darum, dass deren Ergebnisse nur bedingt auf die Domäne Lesen übertragen werden können.

3.4.1 Wirkungskette der schulischen Bildung

Implizit wird immer wieder von einer Wirkungskette ausgegangen, wonach Lehrpersonen Unterrichtsexpertinnen- und -experten sind und über fundiertes Wissen in ihrer Domäne verfügen (Reinisch, 2009). Dieses Wissen wird im Rahmen der Ausbildung erworben (Vogelsang & Reinhold, 2013). Weiter wird in dieser Wirkungskette angenommen, dass die Kompetenzen einer Lehrperson ihre Unterrichtsgestaltung steuern und im Endeffekt zu einer besseren Förderung und zu verbesserten Leistungen der Schülerinnen und Schüler führen (Reinisch, 2009; Vogelsang & Reinhold, 2013). Die folgende Abbildung 9 zeigt diesen angenommenen Weg von der Ausbildung der Lehrperson bis hin zu den Schülerinnen- und Schülerleistungen auf.

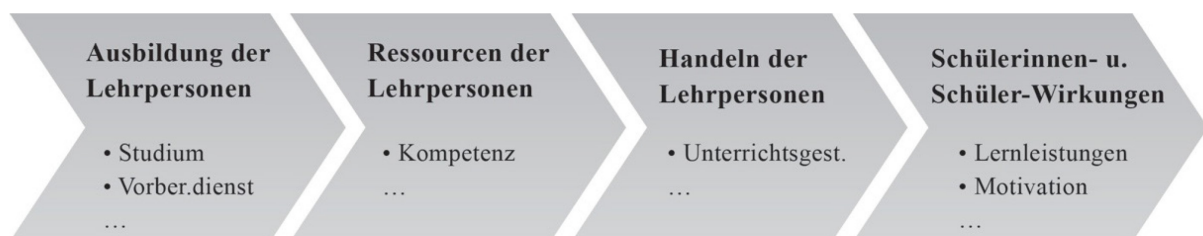


Abbildung 9. Wirkkette der schulischen Bildung aus Vogelsang und Reinhold (2013, S. 319) mit gendergerechter Anpassung durch Jandl (2016, S. 39)

Mit Blick auf empirische Ergebnisse kann heute grundsätzlich davon ausgegangen werden, dass die Kompetenz der Lehrpersonen längerfristig einen relevanten Einfluss auf die Leistungen der Schülerinnen und Schüler hat (u.a. Baumert & Kunter, 2006; Helmke, 2012; Lipowsky, 2006;

Tresch, 2007). Dies gilt zumindest, wenn der Blick auf veränderbare Bedingungsfaktoren wie etwa das Fachwissen und das fachdidaktische Wissen gelenkt wird (König, 2010). Entscheidend ist es, den Fokus auf die Kompetenz von Lehrpersonen *fachspezifisch* anzulegen, da das Fach den eigentlichen Handlungsrahmen von Lehrkräften darstellt und die „Textur des Unterrichts“ beschreibt (Baumert & Kunter, 2006, S. 492). Andererseits wäre der Schluss unzulässig, dass allein das kognitiv abrufbare Wissen darüber entscheidet, ob Kinder gute Leistungen erzielen oder nicht, da weitere Faktoren wie etwa Werthaltungen, motivationale Orientierungen usw. für das unterrichtliche Handeln mitverantwortlich sind (Reinisch, 2009). Zudem muss berücksichtigt werden, dass Wissen keinesfalls automatisch handlungsrelevant wird. Empirische Belege für diese These finden sich u.a. in einer qualitativen Studie zum fachdidaktischen Wissen zum Leseunterricht bei Lehrpersonen bei Scherf (2013). Ergebnisse dieser Studie zeigen, dass handlungsleitendes und kommuniziertes Wissen nicht deckungsgleich sein müssen und dass sich Leseförderlehrpersonen mitunter gar nicht bewusst sind, welches Wissen ihren Handlungen zugrunde liegt. In diesem Zusammenhang ist der Begriff des „trägen Wissens“ bezeichnend, der besagt, dass aufwändig erworbenes Wissen die Praxis oft nicht erreicht (Neuweg, 2014, S. 596). Eine weitere Unsicherheit besteht auf Seiten der Schülerinnen und Schüler. Im oft genannten Angebots-Nutzungsmodell (Helmke, 2012, S. 71) wird aufgezeigt, dass nicht nur das Angebot (Unterricht), sondern auch die Nutzung (Lernaktivitäten der Kinder) darüber entscheidet, ob eine Wirkung erfolgt.

3.4.2 Entwicklung von Fachwissen und fachdidaktischem Wissen

Ausbildung

Im Sinne des Expertiseansatzes wäre anzunehmen, dass hochschulische Lehr-Lernangebote der entscheidende Schlüssel zur Herausbildung professionellen Wissens und Könnens sind und dass die Kompetenzentwicklung ihre Weiterführung im praktischen beruflichen Handeln findet (König, 2010; Kunter et al., 2011b). Zur Annahme, dass sich *fachliches Wissen* im Rahmen der Ausbildung entwickelt, zeigt sich aufgrund empirischer Daten allerdings ein uneinheitliches Bild. Während in der FALKO-Studie (Lindl & Krauss, 2017), der MT-21-Studie (Schmidt et al., 2011) und der COACTIV-Studie (Klusmann, 2011) Leistungszuwächse im Fachwissen über die Studienzeit von Lehramtsstudierenden auf Sekundarstufe I nachgewiesen wurden, konnte dies in anderen Studien nicht festgestellt werden (Länderübergreifende Studie Detuschland/Schweiz: Baer et al., 2007; TEDS-LT: Blömeke, 2013). Etwas einheitlicher sind die Ergebnisse zum *fachdidaktischen Wissen*, das sich im Rahmen der Ausbildung merklich zu entwickeln scheint. In der FALKO-Studie konnte der Zusammenhang allerdings nicht für alle

Fächer bestätigt werden (Lindl & Krauss, 2017). Die nationale TEDS-M-Studie, welche die Wissensentwicklung von Primarlehrpersonen der Schweiz miteinbezieht, zeigt, dass, sich nur im *fachdidaktischen*, nicht aber im fachlichen Bereich positive Veränderungen über die Ausbildungszeit feststellen lassen (Brühwiler, Ramseier & Steinmann, 2015). Der ausbleibende Zuwachs im fachlichen Wissen bei Primarlehrpersonen aus der Schweiz ist möglicherweise dadurch bedingt, dass schulwissensbezogene Inhalte bereits *vor* der Ausbildung erworben und im Rahmen des Studiums nicht weiter ausgebaut werden (ebd.).

Berufserfahrung

Ergebnisse zum Einfluss der Berufserfahrung zeigen ein einheitlicheres Bild in dem Sinne, dass in den meisten Studien keine oder sogar negative Zusammenhänge zwischen Fachwissen und Anzahl Berufsjahre festgestellt wurden (Krauss et al., 2017b; Lindl & Krauss, 2017). Auch zwischen fachdidaktischem Wissen und der Anzahl Berufsjahre konnten sowohl in der FALKO-Studie (Lindl & Krauss, 2017) als auch in der COACTIV-Studie (Brunner et al., 2006) keine signifikanten Korrelationen festgestellt werden.

Auch diesbezügliche Analysen bei der Berufsgruppe der Schulischen Heilpädagoginnen und Heilpädagogen zeigen keinen Effekt auf das fachliche Professionswissen in Mathematik (Jandl, 2016). Folglich ist auch die eher untergeordnete Rolle der Berufserfahrung für den Lernerfolg der Schülerinnen und Schüler wenig erstaunlich (Lipowsky, 2006).

Diese fehlenden Wissenszuwächse im Rahmen der Ausbildung werden u.a. damit begründet, dass vorliegende Instrumente möglicherweise nicht sensitiv auf das Erfahrungswissen von Lehrpersonen reagieren (Lindl & Krauss, 2017). An dieser Vermutung anknüpfend, stellt Neuweg (2014) fest, dass das dem Handeln eingeschriebene implizite Wissen möglicherweise zunehmend quer zu den Kategorien des Ausbildungswissen liegt, und dass das handlungsleitende Wissen von Novizinnen und Novizen zunehmend von kontextsensiblerem Denken und Mustererkennung abgelöst wird.

3.4.3 Einfluss des Fachwissens und fachdidaktischen Wissens auf Schülerinnen- und Schülerleistungen

Zum Zusammenhang zwischen Fachwissen, fachdidaktischem Wissen und Schülerinnen-/Schülerleistungen sind die Ergebnisse der COACTIV-Studie (Kunter et al., 2011a) im Fach Mathematik einschlägig. Bei COACTIV hat sich gezeigt, dass vor allem das fachdidaktische Wissen und weniger das Fachwissen prädiktiv für den Lernfortschritt der Schülerinnen und Schüler zu sein scheint (Lindl & Krauss, 2017, 406). Für das Fachwissen konnten keine vergleichbar signifikanten Effekte nachgewiesen werden (Krauss et al., 2017b). Allerdings kann

das Fachwissen den Erwerb fachdidaktischen Wissens begünstigen (Brunner et al., 2006). Insofern scheint das Fachwissen indirekt Einfluss auf den Leistungszuwachs der Schülerinnen und Schüler zu nehmen (Helmke, 2012) und ist damit eine notwendige, wenn auch nicht hinreichende Bedingung für guten Unterricht (Kunter & Baumert, 2011).

Für den Bereich der Lesekompetenz ist bisher empirisch kaum überprüft, inwieweit die Unterrichtsqualität mit der Leistungsentwicklung der Schülerinnen und Schüler und zudem den Kompetenzen von Lehrpersonen zusammenhängt (McElvany & Schneider, 2009). Studien, die den Zusammenhang zwischen kognitiven Wissensbeständen und Leseleistungen direkt untersuchen, liegen für den deutschsprachigen Raum nach eigenen Recherchen aktuell nicht vor. Dennoch lässt sich aufgrund einzelner deutschsprachiger Studien zumindest indirekt ableiten, dass Fachwissen und didaktische Orientierungen von Lehrpersonen zum erfolgreichen Schriftspracherwerb bzw. Leseerwerb beitragen können. Außerdem lohnt sich der Blick auf eine Studie aus der Domäne Rechtschreiben und auf Studien aus dem angloamerikanischen Raum.

Studien aus dem deutschsprachigen Raum im Bereich Lesen

Als Beleg für die These eines Zusammenhangs zwischen Fachwissen und Leseleistungen können die Ergebnisse der Langzeitstudie von Klicpera und Gasteiger-Klicpera (1993) aufgeführt werden. Die Resultate haben gezeigt, dass die Lese- und Rechtschreibleistungen zwischen den Klassen beträchtlich variieren und dass aufgrund von Unterrichtsbeobachtungen vermutet werden kann, dass neben weiteren Faktoren auch die didaktische Orientierung und die Fachkompetenz von Lehrpersonen entscheidend auf einen erfolgreichen Schriftspracherwerb einwirken. Als weiterer Beleg für die Bedeutung unterrichtlichen Handelns im Hinblick auf die Leseleistung kann der IQB-Ländervergleich (Qualitätsentwicklung im Bildungswesen) beigezogen werden, bei dem an deutschen Schulen moderate Zusammenhänge ($r = .32$) zwischen kompetenzorientierten Aktivitäten der Lehrpersonen und Leseleistungen bei Kindern der neunten Klasse festgestellt wurden (Richter & Lenski, 2015, 728f.). Weiter fanden sich in der Studie von Behrmann und Souvignier (2013) Hinweise darauf, dass sich das Leseverständnis der Schülerinnen und Schüler (fünfte bis siebte Klasse) verbesserte, wenn Lehrpersonen über gute diagnostische Fähigkeiten (Urteilsgenauigkeit) verfügten und zudem überdurchschnittlich viele individuelle Leistungsrückmeldungen gaben. Der Zusammenhang zwischen diagnostischen Fähigkeiten (aufgabenspezifische Treffer) und Lesekompetenzen am Ende der fünften Klasse ließ sich auch bei Karing, Pfof und Artelt (2011) bestätigen.

Ein anderes Bild zeigt sich, wenn das Wissen über *distale* Faktoren erhoben wird. Tiedemann und Billmann-Mahecha (2007) haben untersucht, ob Lehrpersonen, die Deutsch als Fach

studiert hatten, höhere Lesekompetenzen erzielen als fachfremde Lehrpersonen. Die Ergebnisse weisen darauf hin, dass weder das Fachstudium noch die Berufserfahrung einen signifikanten Beitrag zur Aufklärung der Lesekompetenz leisten. Hier muss allerdings berücksichtigt werden, dass das Fachwissen allgemein nur schwache Effekte zeigt, wenn es über distale Merkmale (z.B. besuchte Kurse) erfasst wird (Lipowsky, 2006; Neuweg, 2014). Wie sich die Zusammenhänge von kognitiv abrufbaren Wissensbeständen hinsichtlich Leistungsentwicklung der Schülerinnen und Schüler zeigen, kann aus den erwähnten Studien nicht ermittelt werden.

Studie aus dem deutschsprachigen Raum im Bereich Orthographie

In der Studie von Corvacho del Toro und Thomé (2013) wurden die Auswirkungen des orthographischen Fachwissens von Grundschullehrpersonen auf die Rechtschreibleistung von Schülerinnen und Schülern der zweiten Klasse untersucht. Es hat sich gezeigt, dass das Wissen der Lehrpersonen nicht signifikant mit der mittleren Rechtschreibleistung der Schülerinnen und Schüler zusammenhängt (Corvacho del Toro & Thomé, 2013). Dieses Ergebnis muss allerdings dahingehend präzisiert werden, dass in Klassen, in denen die Lehrperson über weniger fachliches Wissen zum Schriftspracherwerb verfügt, die Anzahl der Rechtschreibfehler der einzelnen Schüler stärker von deren kognitiven Voraussetzungen abhängt, während ein solcher Einfluss der Intelligenz auf die Anzahl der Rechtschreibfehler bei sehr hohem Fachwissen der Lehrperson kaum noch vorhanden war. Insofern scheint sich das Wissen von Lehrpersonen zumindest bei Kindern mit geringen kognitiven Leistungen kompensatorisch auszuwirken. Die Schlussfolgerung, dass besonders schwächere Schülerinnen und Schüler von einer fachlich qualifizierten Lehrperson besonders profitieren würden, wurde allerdings auch kritisiert; ein solcher Schluss könne aus den vorgelegten Daten nicht ohne weiteres gezogen werden (Brügelmann, 2014).

Studien aus dem angloamerikanischen Raum im Bereich Lesen

Aus dem angloamerikanischen Raum liegen Studien vor, welche die Zusammenhänge von Wissen und Leseleistungen von Schülerinnen und Schülern direkt untersuchten. In einer Studie von McCutchen et al. (2002) bei Kindergartenlehrpersonen und Lehrpersonen der ersten und zweiten Klasse wurden Kenntnisse von Lehrpersonen zu angemessener Kinderliteratur und Wissen über strukturelle, linguistische Aspekte der Sprache (u.a. phonologische Bewusstheit) in verschiedenen Tests erfasst und mit der unterrichtlichen Instruktion und den Leseleistungen (Wortlesen, Leseverständnis) der Kinder korreliert. Während bei Kindergartenkindern signifikante Zusammenhänge zwischen dem Wissen der Lehrpersonen, beobachteter Instruktion und Wort-

Leseleistungen festgestellt wurden, blieben diese Zusammenhänge bei Kindern der ersten und zweiten Klasse aus (ebd.). Zu ähnlichen Ergebnissen kommen Carlisle et al. (2009) in einer Studie bei Lehrpersonen der ersten bis dritten Klasse. Da die Lehrpersonen in die drei Kompetenzgruppen „low, middle or high knowledge teachers“ eingeteilt wurden, interessierte hier auch der Zusammenhang in Bezug auf das vorliegende Wissensniveau (ebd., S. 467). In den Ergebnissen der Studie findet sich allerdings kein Zusammenhang des Lehrpersonenwissens mit den Leseverständnisleistungen der Schülerinnen und Schüler. Einzig für die dritte Klasse konnten moderate Unterschiede zwischen den Kontrastgruppen der *high* und der *low knowledge teachers* sowie dem Leseverstehen festgestellt werden (ebd.). Mögliche Erklärung für den ausbleibenden Effekt sehen die Autorinnen und Autoren vor allem im Test selbst, da dieser möglicherweise zu wenig auf die didaktischen Aktivitäten fokussiert und damit nicht das Wissen abgefragt habe, das Lehrpersonen tatsächlich brauchen, um erfolgreichen Leseunterricht durchzuführen. In einer späteren Studie von Carlisle et al. (2011) wurden verschiedene messmethodische Mängel behoben, und die Untersuchung wurde ein weiteres Mal bei Lehrpersonen der ersten bis dritten Klasse durchgeführt. Die erwarteten Effekte blieben auch diesmal aus, und es zeigte sich lediglich ein schwacher Einfluss des Wissens der Lehrperson auf das Leseverständnis am Ende der ersten Klasse (ebd.).

Eine aufschlussreiche Ergänzung zu bisherigen Studien findet sich bei Piasta, McDonald Connor, Fishman und Morrison (2009). In ihrer Untersuchung wurde nicht nur das Wissen der Lehrperson erhoben, auch die Aktivitäten im Klassenzimmer wurden mittels Videobeobachtung aufgezeichnet. Die Untersuchung bei Kindern der ersten Klasse hat verdeutlicht, dass der Lernzuwachs wie angenommen weder allein durch das Wissen der Lehrpersonen noch allein durch die Menge an gezeigter Instruktion von Dekodierstrategien erklärt werden kann (ebd.). Erst das Zusammenspiel von Wissen über Dekodierstrategien und deren Anwendung im Unterricht hat einen signifikanten Lernzuwachs bei Schülerinnen und Schülern bewirkt. Im Weiteren wurde nachgewiesen, dass Lehrpersonen mit hohen Werten im Wissenstest umso bessere Leseleistungen bei den Schülerinnen und Schülern erreichten, je mehr Zeit sie für die direkte Instruktion von Dekodierstrategien verwendet hatten. Demgegenüber erreichten Lehrpersonen mit niedrigen Werten im Wissenstest, die aber viel Zeit mit der Instruktion von Dekodierstrategien verbrachten am Ende sogar niedrigere Wortleseleistungen bei den Schülerinnen und Schülern als zum ersten Testzeitpunkt. Damit zeigt sich, dass falsch angewendetes Wissen sogar „schädliche“ Auswirkungen haben kann (ebd., S. 242).

3.5 Zusammenfassung und Fazit zum Professionswissen von Lehrpersonen

Die Operationalisierung professioneller Kompetenzen bzw. des Professionswissens von Lehrpersonen findet seine theoretische Grundlegung in der Typologisierung von Shulman. Mit Bezug auf Shulman hat sich sowohl in den USA als auch im deutschsprachigen Raum ein reges Interesse an der Messung kognitiver Wissensbestände entwickelt – zunächst im Bereich Mathematik, später auch in anderen Domänen und Fächern. Neben der Untersuchung von „Kompetenz“, die auch affektive Komponenten beinhaltet und zudem stärker handlungsbezogen ist, stand in verschiedenen Studien das kognitiv abrufbare Professionswissen (eine Teilkomponente professioneller Kompetenz) im Fokus. Ein prominent diskutiertes Modell wurde im Rahmen der COACTIV-Studie empirisch überprüft. Das Modell knüpft am Expertiseansatz an und bezieht sich auf ein Verständnis von Professionswissen, das durch in der Ausbildung veränderbare kognitive Wissensbestände gekennzeichnet ist. Das gemessene Professionswissen wird klassischerweise in verschiedene Dimensionen gefasst, wobei sich auch hier die Unterscheidung von Fachwissen, fachdidaktischem Wissen und pädagogischem Wissen in Anlehnung an Shulmans Taxonomie etabliert hat. Insbesondere wurden Anstrengungen im Bereich der Operationalisierung des Fachwissens und des fachdidaktischen Wissens unternommen. Unter anderem wurde im Rahmen der COACTIV-Studie ein Vorschlag präsentiert, der aufgrund seiner empirischen Evidenz wohl zu den einflussreichsten Konzepten gehört. Das zugrunde gelegte Modell hat nicht zuletzt deshalb einen prominenten Platz in der Lehrer- und Lehrerinnenkognitionsforschung, weil sich die operationalisierten Dimensionen auf der Ebene der Leistungsfortschritte der Schülerinnen und Schüler niederschlagen. Insbesondere hat sich das fachdidaktische Wissen von Lehrpersonen als entscheidender Indikator für Leistungsfortschritte erwiesen.

Wichtige Fragen stellen sich in Zusammenhang mit der Dimensionierung und Operationalisierung von Professionswissen, so beispielsweise die Frage nach der Verortung des diagnostischen Wissens. Die Unklarheit, ob es sich hierbei um eine eigenständige Dimension handelt oder ob das diagnostische Wissen durch die Nähe zum fachdidaktischen Wissen eher dieser Dimension zugeordnet werden muss, lässt sich allerdings nicht abschließend beantworten und hängt mit der Art und Weise zusammen, wie die Konstrukte erfasst werden. Eine weitere Frage stellt sich im Zusammenhang mit der mentalen Repräsentation von Professionswissen und inwiefern sich Wissen und Können unterscheiden. Während „Wissen“ eher mit deklarativem Faktenwissen assoziiert ist, beinhaltet das „Können“ bereits Prozeduren und Anwendungen dieses Wissens an konkreten Aufgaben. Eine solche Unterscheidung hat messmethodische Folgen, die sich in den kognitiven Anforderungen von Aufgaben niederschlagen. Aufgaben können z.B.

Anforderungen im Sinne des Erinnerns und Abrufens von Fachbegriffen oder darüber hinaus im Sinne des Verstehens, Analysierens oder Bewertens und Kreierens enthalten.

Mit Blick auf den Stand der Forschung lässt sich zusammenfassen, dass sich die Professionsforschung in den letzten zehn Jahren enorm entwickelt hat. Ein deutlicher Fokus lag vorerst auf naturwissenschaftlichen Fächern und im Bereich Mathematik. Neben konzeptionellen Überlegungen aus diesen Bereichen wurden erste Erkenntnisse aus Studien gewonnen, die inhaltlich stärker auf die Domäne Lesen bezogen werden können. Insbesondere haben sich Studien aus dem Bereich Rechtschreiben als fruchtbar erwiesen. Die Studien waren insofern richtungsweisend, als sie sich ebenfalls auf die oben erwähnten Grundlagen von Shulman bzw. das weiterentwickelte Modell der COACTIV-Studie bezogen. Damit liegen inhaltliche und konzeptionelle Grundlagen vor, die es erlauben, ein anschlussfähiges Modell für den Bereich Lesen zu entwickeln.

4 Professionswissen von Lehrpersonen im Bereich Grundlagen zum Leseerwerb und Diagnostik bei Leseschwierigkeiten

4.1 Ableitung des theoretischen Modells des lesebezogenen Wissens/Könnens

Aus vorangegangenen Kapiteln lassen sich zum einen inhaltliche Dimensionen und deren Facetten ableiten, die den Bereich *Grundlagenwissen* und *diagnostisches Wissen* zum Leseerwerb und zu Leseschwierigkeiten umschreiben. Alle in dieser Arbeit entwickelten Items für das Befragungsinstrument lassen sich durch die entsprechende kognitionspsychologisch orientierte Forschung der letzten Jahre stützen (Kap. 2). Zum anderen können auf dem Hintergrund verschiedener Modelle formale Dimensionen und deren Facetten gebildet werden, die das Fachwissen und das fachdidaktische Können beschreiben. Dieses Wissen setzt sich im vorliegenden Verständnis aus kognitiven *Wissensbeständen* und dem anwendungsbezogenen *Können* zusammen (Kap. 3.3.3). Die Unterscheidung von *Wissen* und *Können* wird nachfolgend noch genauer dargelegt. Das Modell des lesebezogenen Wissens/Könnens stützt sich auf Vorarbeiten von Shulman (1986; 1987) sowie die Weiterentwicklungen seiner Typologisierung des Professionswissens von Lehrpersonen (u.a. Baumert & Kunter, 2011a; Corvacho del Toro, 2013; Moats & Foorman, 2003; Phelps & Schilling, 2004). Den Bezugsrahmen für das Befragungsinstrument bildet somit ein theoretisches Modell bestehend aus vier Feldern und sieben Facetten, die nachfolgend erläutert werden.

Da es sich beim vorliegenden Befragungsinstrument um eine Neukonzeption handelt, wurden zur Beschreibung einzelner Facetten auch Anregungen aus anderen Domänen einbezogen (Kap. 3.3.1). Durch die inhaltliche Nähe von Schreib- und Leseerwerb (Prozessmodelle, Entwicklungsstufen, Struktur der Schriftsprache) konnten auf diese Weise Anregungen für das eigene Befragungsinstrument gefunden werden. Insbesondere aus der Studie von Corvacho del Toro (2013) wurden für das eigene Verfahren entscheidende Hinweise zur Operationalisierung entnommen. Aufgrund der inhaltlichen Neuausrichtung des vorliegenden Projektes konnten bestehende Modelle allerdings nicht unverändert übernommen werden, so dass für die Konzeption des Modells verschiedene Entscheidungen über die Struktur getroffen werden mussten. Diese Entscheidungen werden im Folgenden erläutert und begründet.

1) Zusammenlegung von Fachwissen und fachdidaktischem Können mit Stufen mentaler Repräsentation von Wissen

Eine grundlegende Entscheidung bezüglich der Struktur des Modells wurde hinsichtlich der Zusammenlegung der Dimensionen *Fachwissen* und *fachdidaktisches Wissen/Können* mit Stufen mentaler Repräsentationen von Wissen und Können getroffen. In der vorgelegten Studie

wird auf die Dimensionen *Kennen/Erinnern/Abrufen* und *Verstehen/Analysieren /Anwenden* Bezug genommen. Zudem werden die Beschreibungen von TEDS-M (König, 2012) sowie der Studie TEDS-LT (Blömeke, 2011; Bremerich-Vos et al., 2011) übernommen. Im Modell des lesebezogenen Wissens/Könnens wird der in Kap. 3.3.3.1 diskutierten Annahme gefolgt, dass kognitive Prozesse und Dimensionen des Professionswissens zusammenhängen (siehe Modell Reintjes et al., 2016, S. 62). Das *Fachwissen* rückt dabei in die Nähe des deklarativen Wissens, d.h. es liegt in Anlehnung an die Operationalisierung in den TEDS-Studien in Form von *Kennen/Erinnern/Abrufen* vor. Beim *fachdidaktischen Wissen* dagegen wird meist prozeduralisiertes Wissen im Sinne von *Verstehen/Analysieren/Anwenden* erfasst, das an Problemstellungen gebunden ist und entsprechend durch anwendungsbezogene Items erfasst werden muss. Da das prozeduralisierte Wissen auch als *Können* bezeichnet wird und sich damit vom reinen Wissen abhebt, wird im Folgenden von *Fachwissen* und *fachdidaktischem Können* gesprochen. Die Konsequenzen, die sich hinsichtlich gewählter Itemformate ergeben, werden in Kap. 5.3.2 erläutert.

Eine weitere schlüssige Begründung, die Konstrukte *Fachwissen* und *fachdidaktisches Wissen/Können* mit mentalen Repräsentationen zu überlagern, liefert Neuweg (2014) in seiner differenzierten Analyse zum Begriff des Wissens und Könnens und mit der besagten Brüchigkeit der beiden Dimensionen (Kap. 3.3.2): „Stellt man den Anspruch, dass *Fachwissen* nur wirklich verstanden hat, wer es vermitteln kann, dann wird mit der *fachdidaktischen Kompetenz* im Grunde immer auch zugleich die Tiefe des Verständnisses des *Fachwissens* gemessen“ (ebd., S. 591f.).

2) *Diagnostisches Wissen als inhaltliche Kategorie*

Im hier entwickelten Modell wurde entschieden, das diagnostische Wissen auf formaler Ebene nicht als eigenständige Dimension zu operationalisieren, da sich eine Abgrenzung von der Dimension *fachdidaktisches Wissen/Können* als schwierig erweist (Kap. 3.3.1.3). Im theoretischen Diskurs wird die Nähe der Konstrukte diskutiert und das diagnostische Wissen mehrheitlich dem *fachdidaktischen Wissen* zugeordnet. So findet sich das diagnostische Wissen im Strukturmodell bei COACTIV in der Operationalisierung der *fachdidaktischen Kompetenzfacetten* wieder (Kap. 3.2.2). Im hier entwickelten Modell wurde diagnostisches Wissen dagegen verstanden als „Wissen *über*“ diagnostische Aspekte im Umgang mit LRS. Dieser Überlegung folgend, wurde das diagnostische Wissen im Modell als *Inhaltsdimension* formuliert und nicht auf Ebene der formalen Dimensionen *Fachwissen* und *fachdidaktisches Wissen/Können*. Wie in Kap. 3.3.1.3 dargelegt, wird aber auch hier davon ausgegangen, dass diagnostisches Wissen

domänenspezifisch ist und sich sowohl aus fachlichem Wissen wie aus fachdidaktischem Können speist.

3) Inhaltliche Reichweite des Wissens und Verzicht auf eine eigenständige Dimensionierung von „Spezialwissen von SHP“

Da mit dem in der vorliegenden Arbeit entwickelten Befragungsinstrument neben dem Wissen, das am Ende der Grundausbildung zur Lehrperson erwartet wird, auch vertieftes Wissen, das von Schulischen Heilpädagoginnen und Heilpädagogen im Rahmen einer Zusatzausbildung erworben wird, erfasst werden soll, ist die Frage der inhaltlichen Reichweite zu klären (Kap. 3.3.1.1).

Vor dem Hintergrund der Beschreibung von Fachwissen im Rahmen der COACTIV-Studie stellt sich die Frage, auf welchem Niveau das Professionswissen vorliegen muss bzw. in welchem Rahmen es erworben wird. Im Zusammenhang mit dem Fachwissen wurde in der COACTIV-Studie definiert, dass Professionswissen auf der Ebene eines „tieferen Verständnisses der Fachinhalte des Curriculums der Sekundarstufe“ vorliegen muss. Es kann angenommen werden, dass die abgefragten Wissensbestände im Rahmen der Ausbildung zur Lehrperson erworben werden müssen. In Anlehnung an den in Kap. 3.1.2 beschriebenen kompetenz- und wissensorientierten Expertiseansatz wird auch im vorliegenden Projekt davon ausgegangen, dass Professionswissen im Rahmen der Ausbildung erworben wird und sich über domänenspezifische Anforderungen (operationalisierbar in Dimensionen und Facetten) beschreiben lässt. Im vorliegenden Befragungsinstrument wurde diese Festlegung jedoch dahingehend erweitert, dass auch Items konstruiert wurden, die möglicherweise über das Grundlagenwissen einer Lehrperson am Ende der Grundausbildung hinausgehen und allenfalls nur von SHP gelöst werden können. Auf eine eigenständige theoretische Dimensionierung von „ausbildungsspezifischem Spezialwissen für SHP“ wurde jedoch verzichtet, da die empirischen Grundlagen für eine Vorwegnahme zum Vorliegen einer eigenen Wissensdimension für SHP fehlen. Dennoch wurde die Annahme getroffen, dass zwischen Grundlagenwissen, das für das Unterrichten auf Primarstufe relevant ist und spezialisiertem Wissen der SHP *graduelle* Unterschiede bestehen und dass sich das Wissen von Primarlehrpersonen wiederum substantiell von demjenigen der Laien unterscheidet (vgl. leistungsorientierter Expertiseansatz, Kap. 3.1.2). Auf die Erfassung hochspezialisierten Wissens zu komplexen Fällen von Kindern mit Leseschwächen wurde verzichtet.

4) Überschneidung von Facetten und Zuordnung der Items

Im neu entwickelten Modell wurden auf inhaltlicher Ebene zu den zwei Dimensionen *Grundlagenwissen* und *Diagnostisches Wissen* sieben Facetten unterschieden, die in Kap. 5.3.1 genauer beschrieben sind. Überschneidungen dieser Facetten sind zu erwarten, weshalb der dominanteste Aspekt für die Zuordnung der Items zu den Facetten gewählt wurde (Pissarek & Schilcher, 2017). Auch auf Ebene der kognitiven Anforderungen lassen sich verschiedene Facetten formulieren, wie in Kap. 3.3.3 gezeigt wurde. Um die Komplexität des Modells zu reduzieren, wird darauf verzichtet, die Facetten zu den kognitiven Anforderungen getrennt darzustellen. Die Zuordnung der Items zu den Dimensionen Fachwissen und fachdidaktisches Können geschah in erster Linie vor dem Hintergrund der oben beschriebenen Abgrenzung im Zusammenhang mit der mentalen Repräsentation von Wissen und Können. In Tabelle 3 wird das formale Modell des lesebezogenen Wissens/Könnens dargestellt. Die Verortung der 28 Items der Haupterhebung in den entsprechenden Feldern erfolgt in Kap. 5.3.1.

Tabelle 3: Modell des lesebezogenen Wissens/Könnens

Inhaltsbereich	Kognitive Anforderung	
	Fachwissen Kennen/Erinnern/Abrufen	Fachdidaktisches Wissen/Können Verstehen/Analysieren/Anwenden
Facetten zum Grundlagenwissen		
<i>Facette 1</i> Wissen über linguistische Merkmale der deutschen Schriftsprache und Aspekte der Textschwierigkeit		
<i>Facette 2</i> Wissen über Prozessmodelle des Lesens und Determinanten der Lesekompetenz		
<i>Facette 3</i> Wissen über Stufenmodelle von Leseerwerb / Leseentwicklung inkl. Vorläuferfertigkeiten		
<i>Facette 4</i> Wissen über didaktische Konzepte und Hilfestellungen für den Leseunterricht		
Facetten zum diagnostischen Wissen		
<i>Facette 1</i> Wissen über Symptome von Leseschwierigkeiten		
<i>Facette 2</i> Wissen über Leseverläufe bei unauffällig und schwach lesenden Kindern / Wissen über typische Schwierigkeiten bei Stufenübergängen		
<i>Facette 3</i> Wissen über diagnostische Verfahren und Aufgaben zur Erfassung von Kompetenzen zu Teilbereichen des Lesens		

4.2 Operationalisierung der Inhaltsbereiche Grundlagenwissen und Diagnostisches Wissen

Das im Rahmen des Dissertationsprojektes entwickelte Instrument hat den Anspruch, aktuelles und empirisch abgestütztes Wissen zu Bereichen des Lesens abzufragen. Dafür müssen entsprechende empirisch abgestützte Ergebnisse vorliegen. Um die inhaltliche Validität zu gewährleisten, sollten die zwei Bereiche *Grundlagenwissen* und *Diagnostisches Wissen* durch entsprechende Items möglichst gut abgedeckt werden. Die Herleitung der Items geschieht auf Grundlage der einschlägigen Forschungsarbeiten, die in Kap. 2 dargelegt sind. Darüber hinaus konnten unter Rückbezug auf Operationalisierungen aus dem Bereich Rechtschreiben und Lesen (Kap. 3.3.1) sowie aus Forschungen aus dem englischsprachigen Raum inhaltliche Anhaltspunkte gefunden werden, welche Wissensbestände relevant sind. Zur Absicherung der Bedeutsamkeit erfasster Inhalte werden Normen von Expertinnen und Experten herangezogen, wie später in Kap. 5.3.3.1 genauer erläutert.

Grundlagenwissen

Zum Grundlagenwissen werden vier inhaltliche Facetten beschrieben, die das Spektrum relevanten Wissens, über das Lehrpersonen im Anfangsunterricht bis zur vierten Klasse verfügen sollten, umfassen. Hier werden insbesondere das Wissen über mentale Prozesse beim Lesen, über Zusammenhänge der Teilbereiche (hierarchiehohe- und niedrige Prozesse) des Lesens sowie das Wissen über Entwicklungsschritte (inklusive Vorläuferfertigkeiten beim Lesenlernen) erfasst. Weiter wird erwartet, dass (angehende) Lehrpersonen und SHP über das grundlegende Prinzip des deutschen Schriftsprachsystems (Phonem-Graphem-Korrespondenz) Bescheid wissen und erschwerende und erleichternde Bedingungen, die sich durch textseitige sprachliche Strukturen ergeben, kennen. Schließlich umfasst das Grundlagenwissen auch die Fähigkeit, die Qualität von Hilfestellungen einzuschätzen und zu beurteilen, für welche Schülerinnen und Schüler sie angemessen sind.

Diagnostisches Wissen

Zum diagnostischen Wissen werden drei inhaltliche Facetten gebildet, die das Wissen umschreiben, über das (angehende) Lehrpersonen und/oder SHP verfügen sollten, um die Leseleistungen der Schülerinnen und Schüler angemessen und zuverlässig zu diagnostizieren. Lehrpersonen müssen Kenntnisse darüber haben, wie sich Symptome von Leseschwierigkeiten in verschiedenen Phasen des Leseerwerbs typischerweise äußern, zu welchem Zeitpunkt Risiken diagnostiziert werden können und wie sich eine unauffällige Leseentwicklung in Abgrenzung

zum erschwerten Leseerwerb zeigt. Darin einbezogen ist auch das Wissen über typische Lesefehler bei Stufenübergängen, die auch bei unauffälligen Kindern häufig vorkommen. Zudem lässt sich das diagnostische Wissen über Aufgaben operationalisieren, die Kenntnisse über diagnostische Verfahren und das diagnostische Potenzial von Aufgabenstellungen voraussetzen.

4.3 Operationalisierung von Fachwissen und fachdidaktischem Wissen/Können unter Einbezug der kognitiven Anforderungen

4.3.1 Fachwissen

Mit Bezug auf die theoretischen Darlegungen aus Kap. 3.3.1.1 wird hier unter Fachwissen das vertiefte Hintergrundwissen über die beschriebenen Inhalte *Grundlagenwissen* und *diagnostisches Wissen* im Bereich des Leseerwerbs im Anfangsunterricht der ersten bis vierten Klasse verstanden. Es liegt als akademisches Wissen bzw. Können vor, wie es im Zuge der Grundausbildung zur Lehrperson und der Zusatzausbildung zur SHP erworben wird. Erfahrungsbezogene Erweiterungen dieses Wissens können im hier definierten Fachwissen eingeschlossen sein. Diese Umschreibung stützt sich auf die Definition von Fachwissen im COACTIV-Modell (Baumert & Kunter 2011, 37) und daraus abgeleiteten Operationalisierungen in der Domäne Rechtschreiben (Corvacho del Toro, 2013). Inhaltliche Anregungen für die Domäne Lesen wurden auch aus Studien aus dem amerikanischen Sprachraum sowie aus bestehenden Messinstrumenten im Bereich Lesen gewonnen. In Anlehnung an Shulman (1987) kann das Fachwissen in verschiedenen Facetten umschrieben werden (Kap. 3.2.1) und umfasst mit Bezug auf die inhaltlichen Dimensionen das Folgende:

1) Wissen über Fakten, Umfang und Organisation eines Lerninhaltes

Dazu gehört das Verständnis von grundlegenden linguistischen Merkmalen der deutschen Sprache und von Textschwierigkeit; das Wissen über den Leseprozess und Determinanten der Lesekompetenz auf Wort-, Satz- und Textebene; das Wissen über den Leseerwerb inkl. Vorläuferfertigkeiten; das Wissen über Merkmale leseschwacher Kinder; das Wissen über diagnostische Verfahren zur Erfassung von Leseschwierigkeiten.

2) Fähigkeit, Sachwissen in einen größeren Zusammenhang zu stellen

Dies betrifft das Wissen über Zusammenhänge von hierarchiehöheren und hierarchieniedrigen Leseprozessen und Vorläuferfertigkeiten; das Wissen über Anforderungen von Leseaufgaben in Abhängigkeit von Lesekompetenzen; das Wissen über Hürden beim Lesenlernen im zeitlichen Verlauf.

3) Überblicken des aktuellen Stands der Forschung

Seifert (2015, S. 139) erwähnt diesen Aspekt in diesem Zusammenhang explizit. Auch im Zuge der Beschreibung von Standards (Oser, 1997) wird deutlich, dass Fachwissen theoriegeleitet sein und auf empirisch überprüften Modellen beruhen muss. Damit wird auf übergeordneter Ebene darauf hingewiesen, dass alle Wissensbereiche auf dem aktuellen Stand der Forschung beruhen müssen und dieser bei den Befragten – zumindest implizit – bekannt sein muss. Mit dieser Facette wird u.a. berücksichtigt, dass auch Fehlkonzepte zu einem Fachbereich, die einer empirischen Forschung nicht standhalten, bekannt sein müssen.

Die hier beschriebenen Facetten zum Fachwissen wurden im Modell nicht als einzelne Subkategorien aufgenommen und operationalisiert, sondern dienen hier lediglich der genaueren Beschreibung des Begriffs.

Erfassung des Fachwissens auf der Ebene von „Kennen/Erinnern/Abrufen“

Wie bereits dargelegt, wird das Fachwissen auf der Ebene von Kennen/Erinnern/Abrufen erfasst (Kap. 3.3.3.2). Hier wird Wissen benötigt, das zum Lehrbuchwissen und/oder zum Erfahrungswissen zählt und bei dem es um das korrekte Erkennen und Nennen von Sachverhalten geht. Testaufgaben dieses Typs erfordern von der befragten Person, dass sie zu einem Begriff Phänomene oder Elemente aufzählen kann und Begriffe und Konzepte erkennt (König, 2012). Weiter wird auf dieser Ebene das Faktenwissen abgefragt, etwa das Benennen typischer Merkmale leseschwacher Kinder, ohne dass diese Merkmale in konkreten Situationen bereits erkannt werden müssen. Als Beispiel eines Items auf Ebene des Fachwissens und der entsprechenden kognitiven Anforderung, kann folgendes Item dienen (Abbildung 10).

Die phonologische Bewusstheit spielt für die frühe Leseentwicklung eine wichtige Rolle. Welche der folgenden Punkte gehören zur phonologischen Bewusstheit im engeren und weiteren Sinne? (Mehrere Antworten ankreuzen)		
a	Das Bilden von Reimen.	<input type="checkbox"/>
b	Das Klatschen von Silben bei Wörtern.	<input type="checkbox"/>
c	Das Zuordnen der richtigen Laute zu Buchstaben.	<input type="checkbox"/>
d	Das Analysieren von sprachunabhängigen Klängen und Geräuschen.	<input type="checkbox"/>
e	Das Heraushören von Lauten in Wörtern.	<input type="checkbox"/>
f	Das automatisierte Benennen von Objekten (Bilder von Gegenständen)	<input type="checkbox"/>

Abbildung 10. Beispiel-Item 8 Fachwissen (Kennen/Erinnern/Abrufen)

4.3.2 Fachdidaktisches Wissen/Können

Diese Dimension ist in der vorliegenden Arbeit über drei Facetten operationalisiert, die an Shulman (1986) Typologisierung und Weiterentwicklungen davon anknüpfen. Insbesondere

werden Ausführungen zu den drei Facetten aus der COACTIV-Studie (Baumert & Kunter, 2011a) und die Operationalisierung in der Domäne Rechtschreiben (Corvacho del Toro, 2013; Riegler & Wiprächtiger-Geppert, 2016) einbezogen. Weitere Anregungen aus der Domäne Lesen stammen aus der Studie von Rutsch (2016) sowie aus Studien aus dem amerikanischen Sprachraum (Phelps & Schilling, 2004) (Kap. 3.3.1.2).

Mit Bezug auf diese theoretischen Vorarbeiten wird hier unter fachdidaktischem Wissen/Können die Anwendung von Grundlagenwissen und diagnostischem Wissen zu spezifischen lesebezogenen Anforderungen im Anfangsunterricht der ersten bis vierten Klasse verstanden. Items zu dieser Dimension erfassen, inwiefern es (angehenden) Lehrpersonen und SHP gelingt, Aufgabenschwierigkeiten einzuordnen und geeignete lernförderliche Hilfestellungen für Schülerinnen und Schüler zu identifizieren oder zu entwickeln. Darüber hinaus umfasst dieses Wissen das Analysieren von typischen Lesefehlern in konkreten Situationen und das korrekte Einschätzen von Lesekompetenzen. Damit ist ein zentraler Aspekt diagnostischen Wissens angesprochen. Weiter wird hier unter fachdidaktischem Können auch die Anwendung von linguistischem Wissen (Graphem-Phonem-Korrespondenz) bei praktischen Aufgaben verstanden. Die Dimension fachdidaktisches Wissen/Können wird unter Rückbezug auf oben genannte theoretische Arbeiten in der vorliegenden Arbeit über drei Facetten operationalisiert:

1) Aufgabenbezogenes Wissen

Wissen über die Qualität von Aufgaben und Auswahl geeigneter Aufgaben in bestimmten Phasen der Leseentwicklung; Wissen über das diagnostische Potenzial sowie die kognitiven Anforderungen und Wissensvoraussetzungen von Aufgaben.

2) Wissen über Schülerinnen- und Schülerkognitionen und Fehlerwissen

Wissen über Vorstellungen von Kindern sowie über typische Fehler oder Konzepte, die Kinder aufgrund ihres Alters mitbringen; Wissen über Repräsentationen und Erklärungen einschließlich dem schnellen Erkennen von Fehlern.

3) Wissen über den Lernprozess fördernde Hilfestellungen

Wissen über geeignete Hilfestellungen zur Überwindung von lesebezogenen Schwierigkeiten; Wissen über geeignete Fördermethoden.

Die hier beschriebenen Facetten zum fachdidaktischen Wissen/Können wurden im Modell, analog zum Fachwissen nicht als einzelne Subkategorien aufgenommen und operationalisiert, sondern dienen auch hier lediglich der genaueren Beschreibung der Dimension.

Erfassung des fachdidaktischen Wissens auf der Ebene von „Verstehen/ Analysieren/ Anwenden“

Neben dem Erinnern von Wissen geht es beim Verstehen/Analysieren/Anwenden darüber hinaus um die Verbindung dieses Wissens mit einer Problemstellung. Beim *Können* im Sinne von prozeduralem Wissen sind Operationen und Aufgaben gemeint, bei denen die Probandinnen und Probanden selbst etwas tun müssen. Testaufgaben, die diese Qualität von Wissen erfassen, erfordern von den Testpersonen, dass sie Sachverhalte in bestimmten Situationen analysieren, erkennen, interpretieren und einordnen können (König, 2012). Studierende müssen auf dieser Stufe gewissermaßen „beweisen“, dass sie ihr erworbenes Fachwissen nicht nur aus der Erinnerung abrufen, sondern dieses Wissen bei Aufgabenstellungen umsetzen können und somit einen Sachverhalt auch *verstanden* haben. Als Beispiel eines Items auf Ebene des fachdidaktischen Könnens und der entsprechenden kognitiven Anforderung kann folgendes Item dienen (Abbildung 11).

Sätze und Texte können auf der Textoberfläche unterschiedlich schwierig zu lesen und zu verstehen sein. Ordnen Sie die folgenden 5 Textausschnitte nach Schwierigkeit. Die leichteste Aufgabe erhält die 1, die schwierigste die 5.	
	Schwierigkeit
a	Die Kätzchen spielen im Eingang und beobachten einen Spatz. Die Kätzchen verschrecken plötzlich und rennen davon.
b	Der Hund ist im Haus und hat nun Durst. Der Hund will nun Wasser haben.
c	Die Grossmutter erzählt am Wochenende oft viele Geschichten und strickt Strümpfe dabei. Sie legt diese danach jeweils in eine schöne, grosse Truhe.
d	Die Kinder schreiben tolle Geschichten und singen frohe Weihnachtslieder in der Schule. Die Kinder haben einen Stern gemacht und in Papier eingepackt.
e	Die Kinder warten vor dem Gartentor und haben Hunger. Die Kinder dürfen nun Käsebrot essen.

Abbildung 11. Beispiel-Item 26 fachdidaktisches Wissen/Können (Verstehen/Analysieren/Anwenden)

5 Darstellung der Untersuchung und Entwicklung des Befragungsinstruments

5.1 Ableitung des Forschungsanliegens und der Forschungsziele

Aus den vorangehenden Kapiteln wird deutlich, dass das Professionswissen von Lehrpersonen bislang vor allem in den Bereichen Mathematik und Naturwissenschaften (Kunter et al., 2011a) untersucht wurde. Inzwischen liegen zwar Studien zu anderen Fächern vor (z.B. FALKO (Krauss et al., 2017a), doch es zeigt sich, dass sich nur wenige Untersuchungen mit der Kompetenz von Lehrpersonen im Deutschunterricht befassen (Blömeke et al., 2011). Ausnahmen bilden die Studien PlanvoLL-D (Bremerich-Vos et al., 2016) und FALKO-D (Pissarek & Schilcher, 2017) sowie für die Domäne Schreiben u.a. die Studien von Corvacho del Toro (2013). Studien, die explizit auf die quantitative Kompetenzmessung bei Lehrpersonen im Bereich Lesen abzielen, finden sich vor allem im US-amerikanischen Raum (u.a. Carlisle et al., 2011; Moats & Foorman, 2003; Piasta et al., 2009). Für den deutschsprachigen Raum bilden in diesem Bereich die Studien von Rutsch (2016) sowie Schmidt und Schabmann (2016) die Ausnahmen. Diese Studien weichen allerdings bezüglich Stufe und/oder Modellierung des Professionswissens vom Vorhaben in der vorliegenden Arbeit ab. Eine zeitgleich zu diesem Dissertationsprojekt durchgeführte Studie aus Deutschland (Hanke et al., 2019) ist hinsichtlich Stufe (Primarstufe) und Konzeption ähnlich. Ein auf die Ausbildungssituation von Lehrpersonen an Pädagogischen Hochschulen der Schweiz abgestimmtes Instrument fehlt aber derzeit noch. Zur Schließung dieser Forschungslücke soll die vorliegende Studie einen Beitrag leisten. Hauptanliegen der Studie ist die Entwicklung eines Befragungsinstruments, das das Wissen und Können von Studierenden mit verschiedenen Ausbildungs- und Berufserfahrungsvoraussetzungen im Bereich Anfangsleseunterricht erfasst. Um den späteren Verwendungszweck des Befragungsinstruments zu erfüllen – nämlich Lehrpersonen mit unterschiedlichem Ausbildungshintergrund – vergleichen zu können, wurden für die Entwicklung des Instruments Studierende am Ende der Grundausbildung ohne Berufserfahrung, Lehrpersonen mit Berufserfahrung sowie Studierende der Schulischen Heilpädagogik (SHP) einbezogen. Das Testinstrument basiert auf dem aktuellen Forschungsstand lesedidaktischer Konzepte für die erste bis vierte Klasse und bezieht die besondere Situation von Kindern mit Leseschwierigkeiten mit ein (Kap. 2).

Die theoretische Konzipierung des Instrumentes beruht auf der Grundlage des Modells zum Professionswissen von Baumert & Kunter (2006). Das Modell berücksichtigt die auf Shulman (1987) zurückgehenden Wissensbereiche *Fachwissen* und *fachdidaktisches Wissen*. Diese Unterscheidung lag auch der hier vorgestellten Vorgehensweise zugrunde.

Fachwissen und fachdidaktisches Wissen/Können wurden zu folgenden zwei Inhalten erfasst: *Grundlagen zum Leseerwerb und Lesekompetenz* und *Diagnostik bei Leseschwierigkeiten*.

Anregungen für inhaltliche Spezifizierungen und das Bilden von Facetten gaben verschiedene in Kap. 3.3 erwähnte Studien. Diese Grundlagen wurden auf den Bereich des Lesens übertragen und erweitert. Für das selbst entwickelte Befragungsinstrument erfolgte eine Eingrenzung auf Teilbereiche des Lesens der ersten bis vierten Klasse. Das entwickelte Instrument fokussiert ausschließlich auf den Bereich des Lesen/lernens, also auf Aspekte, die mit dem „learning to read“, und weniger mit dem „reading to learn“ zusammenhängen (Rosebrock & Nix, 2012, S. 75). Somit fließen Aspekte der Anschlusskommunikation, Lesesozialisation und Lesen im Sinne literarischer Bildung und als Kulturgut (Hurrelmann, 1994; Rosebrock & Nix, 2012) nicht in das Befragungsinstrument ein. Ausgehend davon ergab sich ein theoriebasiertes Modell, auf dessen Grundlage das Untersuchungsinstrument entwickelt wurde und welches zur Einhaltung der inhaltlichen Validität herangezogen wurde.

Methodische Einordnung

Da Projekt ist in der Tradition der Professionsforschung und genauer der Lehrerkognitionsforschung anzusiedeln, was sich u.a. im methodischen Vorgehen abzeichnet. Bei dieser Forschungsrichtung geht es darum, empirisch gehaltvolle Aussagen über die (Handlungs-) Expertise von Lehrpersonen zu machen, die über die Untersuchung des Denkens von Lehrpersonen gewonnen werden (Reinisch, 2009). Wie im Forschungsüberblick (Kap. 1) gezeigt, sind im Bereich des Lesens erst wenige Studien zu finden, die dem Anspruch genügen, kognitive Aspekte von Wissen und Können objektiv *quantitativ* zu erfassen und in diesem Sinne das Kriterium der Objektivität erfüllen (Döring & Bortz, 2016). Ebenso besteht ein Bedarf an Instrumenten, die auch ökonomisch für Untersuchungen mit großen Stichproben eingesetzt werden können (Rupp & Bonholt, 2006).

In der vorgestellten Untersuchung wurden überwiegend quantitative Methoden angewendet. Qualitative Verfahren (Expertinnen- und Expertenbefragung und kognitives Pretesting) kamen im Lauf der Entwicklung der Items nur zu einem kleinen Teil zum Zuge (Kap. 5.3.3). In vorliegender Studie wurde ein *neues* Befragungsinstrument entwickelt; in diesem Sinne wurden keine Hypothesen geprüft, sondern vielmehr Forschungsanliegen und entsprechende Ziele formuliert. Einzig im Rahmen der Validierung des Befragungsinstruments wurde der Sachverhalt ausgenutzt, dass drei Substichproben mit unterschiedlichem Hintergrund bezüglich Ausbildungs- und Berufserfahrung und zudem eine Laien-Kontraststichprobe einbezogen wurde (Krauss et al., 2011, 152; Rutsch, 2016). Es wurde angenommen, dass sich die Ausprägung der erfassten Fähigkeiten in Abhängigkeit von der Berufs- und Ausbildungsbiografie zeigen, wie dies in zahlreichen anderen Studien ebenfalls angenommen wurde (Zusammenfassend bei

Jandl, 2016). Da Studien der *Testerprobung* und *Testanwendung*, bei denen Ergebnisse (z.B. Gruppenunterschiede) schließlich auch verwertet werden, strikte getrennt werden sollten (Lienert & Raatz, 1998), ist das Ergebnis zu Gruppenunterschieden lediglich Grundlage, um das Instrument hinsichtlich seiner Ausbildungssensitivität zu beurteilen und kann nicht als Bestätigung einer Hypothese interpretiert werden.

Forschungsanliegen 1: Entwicklung eines ökonomischen und objektiven Befragungsinstruments

Mit dem Projekt wird das Ziel verfolgt, ein für weitere Forschung anwendbares Befragungsinstrument zu entwickeln, das ökonomisch als Gruppenverfahren eingesetzt werden kann und eine fundierte Einschätzung bezüglich der diagnostischen Fähigkeiten und des Grundlagenwissens von Lehrpersonen und SHP im Bereich Lesen und Leseschwierigkeiten erlaubt. Dabei stellt der Einbezug von SHP eine Besonderheit dar, einerseits weil sich den SHP aufgrund ihrer Ausbildung bei der Prävention und Unterstützung leseschwacher Kinder eine besondere Aufgabe stellt (Hartmann, 2007), andererseits weil für diese Berufsgruppe ein Mangel an quantitativen Forschungsergebnissen zum Professionswissen vorliegt (Moser & Kropp, 2014). Auf einer allgemeinen Ebene lassen sich im Folgenden zwei Forschungsziele bzw. Ansprüche an das Befragungsinstrument formulieren:

Ziel 1: Das Befragungsinstrument soll in seiner Durchführung und Auswertung objektiv sein.

Ziel 2: Das Befragungsinstrument soll in seiner Durchführung ökonomisch sein und sich als Gruppenverfahren eignen für Lehrpersonen mit unterschiedlichem Vorwissen, bedingt durch Ausbildung und/oder Berufserfahrung.

Forschungsanliegen 2: Skalierung eines Befragungsinstruments und Beurteilung der Testgütekriterien

Die einzelnen Schritte zur Entwicklung des Befragungsinstruments werden in Kap. 5.3 beschrieben. Ein zentrales Anliegen besteht darin, das Instrument hinsichtlich seiner psychometrischen Eigenschaften zu prüfen. Dies kann einerseits auf der Grundlage der *Klassischen Testtheorie* (KTT) geschehen, andererseits auf der Grundlage der *Item-Response-Theorie* (IRT). Über viele Jahre hinweg war die Klassische Testtheorie der vorherrschende Ansatz der Testkonstruktion (Eid & Schmidt, 2014), dies hat sich aber in den vergangenen Jahren verändert. Insbesondere werden im Rahmen der Professionsforschung und den dabei entwickelten Instrumenten vermehrt Skalierungen auf der Grundlage von IRT-Modellen vorgenommen (TEDS-LTBlömeke et al., 2013a; COACTIV-Studie Kunter et al., 2011a). Mit der Anwendung von

IRT-Modellen werden Schwächen der KTT teilweise überwunden, wie in Kap. 6.1.2 genauer dargelegt wird. In der vorliegenden Untersuchung wurde entschieden, die Skalierung auf Grundlage beider Testtheorien durchzuführen, da sich diese grundsätzlich gut ergänzen (Seifert, 2015).

Die Besonderheiten der IRT und insbesondere des Rasch-Modells werden in Kap. 6.1.3 genauer eingeführt. Für beide Testtheorien werden zentrale Testgütekriterien formuliert, aufgrund derer die Qualität des entwickelten Befragungsinstruments beurteilt werden soll. Daraus leitet sich das dritte Ziel der Untersuchung ab:

Ziel 3: Das entwickelte Befragungsinstrument soll unter Anwendung der KTT und der IRT skaliert und auf seine psychometrischen Eigenschaften hin überprüft werden. Namentlich sind dies Itemschwierigkeit, Reliabilität und Trennschärfe.

Forschungsanliegen 3: Beurteilung der latenten Struktur des Professionswissens von Lehrpersonen und SHP

In Kap. 4 wurde das zugrunde gelegte Modell für die Entwicklung des Befragungsinstruments dargelegt. Die Dimensionen *Fachwissen* und *fachdidaktisches Wissen/Können* leiten sich dabei aus Kompetenzstrukturmodellen empirischer Forschung ab. Inwiefern sich diese – und zwei weitere inhaltliche Dimensionen – auch in der vorliegenden Untersuchung zeigen oder ob sich alternative latente Strukturen ermitteln lassen, soll über Modellvergleiche sowie eine exploratorische Faktorenanalyse erfasst werden. Damit ergeben sich zwei weitere Ziele:

Ziel 4: Das entwickelte Befragungsinstrument soll hinsichtlich der postulierten Dimensionen des theoretisch abgeleiteten Modelles zum Professionswissen überprüft werden. Dies betrifft die beiden Inhaltsgebiete *Grundlagen zum Leseerwerb* und *Diagnostik bei Leseschwierigkeiten* sowie die beiden kognitiven Anforderungen *Fachwissen* und *fachdidaktisches Wissen/Können*.

Ziel 5: Beim entwickelten Befragungsinstrument soll geprüft werden, ob andere Dimensionen als die theoretisch postulierten vorliegen oder ob das Professionswissen als eindimensionales Konstrukt strukturiert ist.

Forschungsanliegen 4: Validierung des Befragungsinstruments

Zur Validierung des Befragungsinstruments wurden qualitative und quantitative Verfahren angewendet. Ziel der qualitativen Validierung war es sicherzustellen, dass das Instrument theoretisch und praktisch relevante Items enthält, die das beabsichtigte Konstrukt angemessen erfassen. In Anlehnung an das Vorgehen bei FALKO-D (Pissarek & Schilcher, 2017) wurden zu diesem Zweck Items durch Expertinnen und Experten beurteilt und anschließend in einem kognitiven Pretesting-Verfahren durch potenzielle Probandinnen und Probanden beantwortet. Im Anschluss daran wurden ungeeignete Items eliminiert (Kap. 5.3.3).

Ziel 6: Das Befragungsinstrument soll inhaltstvalide sein, indem theoretisch sowie praktisch relevante Inhalte erfasst werden, die durch Expertinnen und Experten als solche beurteilt werden.

Im Rahmen des quantitativen Validierungsprozesses wurde zunächst analysiert, inwiefern beim Befragungsinstrument Messinvarianz vorliegt, d.h., ob in verschiedenen Substichproben dasselbe Konstrukt gemessen wird. In anderen Studien wird entlang eines relevanten Teilungskriteriums wie z.B. Sprache (HarmoS-Studie Ramseier, Moser, Moreau & Antonietti, 2008), Geschlecht (Lihuan & Anthony, 2018), Land (Schindler, Moser Opitz, Cadonau-Bieler & Ritterfeld, 2019) ermittelt, ob Items in den Subgruppen variieren und überproportional schwieriger oder leichter sind (Blömeke et al., 2009). Treten solche Unterschiede auf, kann dies als Indikator dafür gelten, dass für das Lösen entsprechender Items ein anderes Professionswissen als das zentral modellierte notwendig ist. In der vorliegenden Studie ist das Teilungskriterium die Berufsbiografie (Ausbildung/Berufserfahrung). Damit lässt sich untersuchen, welche Items in den drei Substichproben 1) Lehrpersonen ohne Berufserfahrung am Ende des Studiums 2) Lehrpersonen mit Berufserfahrung und 3) SHP am Ende des Studiums substanziell unterschiedlich gelöst wurden. Im vorliegenden Fall gab die IRT-basierte Differenzielle Itemfunktion-Ansatzes (DIF) Aufschluss über das Vorliegen von Messinvarianz (Schwab & Helm, 2015). Damit verbunden lässt sich folgender Anspruch ableiten:

Ziel 7: Mit dem Befragungsinstrument soll in verschiedenen, nach Berufsbiografie gebildeten Substichproben, dasselbe Konstrukt erfasst werden.

Zur Beurteilung der Validität wurden drei weitere Verfahren verwendet. Zum einen wurde die Bildung von drei Substichproben mit unterschiedlichem Ausbildungshintergrund ausgenutzt, um das Professionswissen in verschiedenen Ausprägungen zu messen. Die Vermutung, dass sich Professionswissen ausbildungsspezifisch zeigt und sich je nach Berufs- bzw. Ausbildungs-

biografie unterschiedlich darstellt, wird u.a. bei Jandl (2016) für das Fach Mathematik bestätigt. Allerdings wirkt sich die Berufserfahrung von SHP in den Auswertungen von Jandl hypothesenkonform *nicht* auf das mathematikspezifische Professionswissen aus. SHP mit abgeschlossener Ausbildung in Schulischer Heilpädagogik weisen sogar ein geringeres mathematikspezifisches Professionswissen auf als Studierende in Schulischer Heilpädagogik am Ende ihrer Ausbildung (Jandl, 2016; Jandl & Moser Opitz, 2017). Dieser fehlende Zusammenhang zwischen der Anzahl der Jahre mit Berufserfahrung und dem gemessenen Professionswissen zeigt sich in verschiedenen Fächern auch in der FALKO-Studie (Pissarek & Schilcher, 2017). Für den Bereich Lesen wurden bei Rutsch (2016) entsprechende Hypothesen geprüft. Neben dem Erwerb von lesedidaktischem Wissen, das im Rahmen der Ausbildung erworben wird, schlägt sich auch die *spezifisch lesedidaktische* Lehrerfahrung auf die Testleistung nieder (Rutsch, 2016). Die Forschungslage zum Einfluss spezifischer Lerngelegenheiten und eigener Berufserfahrung hinsichtlich Wissen und Können im Bereich Lesen ist allerdings noch dünn (Rutsch, Schmitt & Dörfler, 2018b).

Der querschnittliche Wissensanstieg der drei Substichproben dieser Arbeit könnte als Indikator gelten, dass das entwickelte Instrument *ausbildungssensitiv* misst. Dabei wird dieses Ergebnis nicht als Bestätigung der Hypothese eines querschnittlichen Wissensanstiegs interpretiert, sondern als Indikator für die Ausbildungssensitivität und die Validität des Instruments gewertet (Krauss et al., 2011).

Ziel 8: Mit dem Befragungsinstrument sollen Wissensunterschiede zwischen Lehrpersonen mit verschiedener Berufsbiografie (Ausbildung/Berufserfahrung) sensitiv erfasst werden.

Ein weiterer Beleg für die Validität des Instruments ergibt sich aus der Erwartung, dass ein deutlicher Wissensunterschied zwischen Laien und Studienteilnehmenden mit ausbildungsspezifischem Professionswissen besteht. Dieser Kontrastgruppenvergleich (teilweise auch als Kontrollgruppe bezeichnet) wurde auch in anderen Studien im Zuge der Validierung von Instrumenten durchgeführt (Krauss et al., 2011; Rutsch, 2016). Allerdings wurde bei COACTIV keine Laienstichprobe als Kontraststichprobe gewählt, sondern es wurden Kompetenzunterschiede im Vergleich mit Studienteilnehmenden verwandter Professionen gemessen (Krauss et al., 2011, 153). In der vorliegenden Studie wurde eine weniger „konservative“ Strategie gewählt, indem Unterschiede im Professionswissen ermittelt wurden zwischen den drei genannten Substichproben und Lehrpersonen (Laien), die ganz am Anfang der Ausbildung standen und keinerlei Vorwissen im Bereich Leseentwicklung, Lesedidaktik und Umgang mit Leseschwierigkeiten mitbrachten. Damit wird als weiterer Aspekt der Validität geprüft, ob das Instrument

tatsächlich Professionswissen misst und nicht beiläufig erworbenes Wissen, das auch ohne entsprechende Ausbildung vorliegt.

Ziel 9: Mit dem Befragungsinstrument soll lesespezifisches Professionswissen erfasst werden, das sich deutlich von allgemeinen Wissensbeständen von Laien abgrenzt.

5.2 Untersuchungsplan

Die Erstellung des Befragungsinstrumentes war auf einen Zeitraum von vier Jahren angelegt und gliedert sich in verschiedene Etappen: eine Voruntersuchung und eine Hauptstudie sowie Entwicklungs- und Überarbeitungsphasen. Der Untersuchungsplan mit den einzelnen Arbeitsschritten und Kapitelbezügen wird in Abbildung 12 vorgestellt. Eine genaue Beschreibung der Entwicklungsschritte und der durchgeführten Methoden zur Erreichung der formulierten Forschungsziele findet sich in den nachfolgenden Kapiteln.

Entwicklung Entwurf des Befragungsinstrumentes und erste qualitative Validierung	Kapitel
Entwicklung von 56 Items auf Grundlage des theoretischen Modells Entwicklung des Kurzfragebogens zur Berufsbiografie der Probandinnen und Probanden Entwicklung des vorläufigen Lösungsschlüssels Validierung durch Expertinnen und Experten ($n = 4$) Kognitives Pretesting ($n = 3$) Reduktion auf 24 Items	Kap. 5.2 Kap. 5.4 Kap. 5.3.4 Kap. 5.3.3.1 Kap. 5.3.3.2
Voruntersuchung Stichprobe $N = 203$	
Erprobung des Entwurfs an drei Pädagogischen Hochschulen und Ausbildungsinstitutionen für Schulische Heilpädagogik der Schweiz Quantitative Analyse des Instrumentes und Itemselektion Item- und Skalenanalysen Dimensionalität Validität Reduktion auf 15 Items → 9 Items werden aufgrund kritischer Testgütekriterien verworfen	Kap. 7.1 Kap. 7.1.1 Kap. 7.1.2 Kap. 7.1.3
Überarbeitung Entwicklung neuer Items und zweite qualitative Validierung	
Überarbeitung bestehender Items und Entwicklung von 46 Items inkl. Lösungsschlüssel Validierung durch Expertinnen und Experten ($n = 5$) Kognitives Pretesting ($n = 6$) Prüfung der Lösbarkeit durch Expertinnen und Experten ($n = 2$) Reduktion auf 28 Items	Kap. 5.2 Kap. 5.3.3.1 Kap. 5.3.3.2
Hauptstudie Stichprobe $N = 507$	
Einsatz des Befragungsinstrumentes an fünf Pädagogischen Hochschulen und Ausbildungsinstitutionen für Heilpädagogik der Schweiz Quantitative Analyse des Instrumentes und Itemselektion Item- und Skalenanalysen → Reduktion auf 18 Items Distraktorenanalyse Differenzielle Itemfunktionsanalyse Dimensionalität Validität	Kap. 7.2 Kap. 7.2.1 Kap. 7.2.2 Kap. 7.2.3 Kap. 7.2.4 Kap. 7.2.5

Abbildung 12. Untersuchungsplan und Arbeitsschritte

Voruntersuchung

In der Voruntersuchung wurde auf der Grundlage des theoriebasierten Modells, das in Kap. 4.1 beschrieben ist, 56 Items entwickelt und in einem mehrschrittigen Prozess reduziert (siehe Vorgehen bei Kirchhoff, 2017). Für die inhaltlich und formal begründete Reduktion wurde sowohl eine Validierung durch Expertinnen und Experten (Kap. 5.3.3.1) als auch ein kognitives Pretesting (Kap. 5.3.3.2) durchgeführt. Im Anschluss an diese inhaltliche Validierung wurde ein Entwurf des Befragungsinstrumentes mit 24 Items bei einer Stichprobe von gesamthaft 203 Studienteilnehmenden erprobt. Die Zusammensetzung der Stichprobe erfolgte dabei nach denselben Kriterien wie später in der Hauptuntersuchung (Kap. 5.4.1) und war darauf angelegt, ein breites Wissensspektrum zu erfassen. Für die Voruntersuchung wurden insgesamt 181 Studierende von drei Pädagogischen Hochschulen und Ausbildungsinstitutionen für Schulische Heilpädagogik einbezogen. Die Stichprobe setzte sich wie folgt zusammen: (1) *Primarlehrpersonen ohne Berufserfahrung am Ende des Studiums* ($n = 66$), (2) *Primarlehrpersonen mit Berufserfahrung* ($n = 61$), (3) *SHP mit Berufserfahrung am Ende des Studiums* ($n = 54$). Die *Kontraststichprobe* bestand aus 22 Laien.

Ziel der Voruntersuchung war einerseits das Eruiere geeigneter Items und Itemformate. Andererseits wurden Aspekte der Durchführbarkeit (zeitliche Belastung, Klarheit der Instruktion, Verständlichkeit der Items) beachtet und daraus entsprechende Schlüsse gezogen. Die statistischen Ergebnisse der Voruntersuchung finden sich in Kap. 7.1.

Nach Überarbeitung des Instrumentes, bei dem 15 theoretisch und empirisch belastbare Items beibehalten werden konnten und 31 Items neu entwickelt werden mussten, lagen für die Hauptuntersuchung 46 Items vor. Diese wurden analog zur Voruntersuchung in einem mehrstufigen qualitativen Validierungs-Prozess beurteilt und schließlich auf 28 Items reduziert.

5.3 Entwicklung Befragungsinstrumentes zum lesebezogenen Wissen/Können

5.3.1 Beschreibung des Befragungsinstrumentes

Grundlage für die Entwicklung des Befragungsinstrumentes war ein theoretisches Strukturmodell, dessen Herleitung und Konzeptualisierung in Kap. 4 ausführlich beschrieben ist. Im Anschluss an diesen ersten Schritt zur Operationalisierung spezifischer Professionalität im Bereich Lesen galt es, dieses Konstrukt über geeignete Indikatoren und Items empirisch messbar zu machen (Böhm-Kasper & Weishaupt, 2008). Im Sinne der *operationalen Merkmalsdefinitionen* („Ist das Item Teil der interessierenden Gesamtheit möglicher Items“) und der *theoretischen Merkmalsdefinitionen* („Kann das theoretische Konstrukt Unterschiede in den beobachteten

Antworten erklären“) (Seifert, 2015, S. 151) wurden durch die Orientierung am Modell des lesebezogenen Wissens/Könnens Aspekte der Inhaltsvalidität berücksichtigt.

Der Prozess der Itementwicklung schloss immer mehrere Fragen mit ein. Zum einen mussten Items entwickelt werden, welche die *inhaltlichen Facetten* (Grundlagen zum Leseerwerb und Diagnostik bei LRS) möglichst umfassend abbilden und deren Bedeutung sich theoretisch belegen lässt. Bei der Konstruktion der Items wurde auf aktuelle Quellen im Bereich Lesen zurückgegriffen – und zwar vorwiegend auf solche, die ihre Aussagen auch auf empirische Forschungsergebnisse stützen (Pissarek & Schilcher, 2017). Obschon in der Literatur bisweilen davor gewarnt wird zu „akademische“ Fragen zu stellen und empfohlen wird, auf Lehrbücher als Inspirationsquelle zu verzichten und diese lediglich zur Absicherung der fachlichen Richtigkeit einzubeziehen (Krebs, 2004, S. 4), wurde in der vorliegenden Arbeit für die Konstruktion der Items „Lehrbuchwissen“ einbezogen. Weiter mussten Kriterien gefunden werden, die eine Kategorisierung der Items nach *kognitiven Anforderungen* (Fachwissen und fachdidaktisches Wissen/Können) zuließ. In der folgenden Tabelle 4 sind die 28 Items der Hauptuntersuchung im Modell des lesebezogenen Wissens/Könnens verortet.

Tabelle 4: Verortung der Items im Modell des lesebezogenen Wissens/Könnens

Inhaltsbereich	Kognitive Anforderung			
	Item Nr.	Fachwissen Kennen/Erinnern/Abrufen	Item Nr.	Fachdidaktisches Wissen/Können Verstehen/Analysieren/Anwenden
<i>Facette 1</i>				
Wissen über linguistische Merkmale der deutschen Schriftsprache und Aspekte der Textschwierigkeit	28	Was ist der Lix und über welche Textmerkmale wird er berechnet	26 27	Ordnen der Satzschwierigkeit Phonemanalyse von 6 Begriffen
<i>Facette 2</i>				
Wissen über Prozessmodelle des Lesens und Determinanten der Lesekompetenz	9 10 19	Zusammenhang Leseflüssigkeit und Leseverständnis Zusammenhang Hörverständnis Leseverständnis Mindestanteil an korrekt gelesenen Wörtern für unabhängiges Lesen		
<i>Facette 3</i>				
Wissen über Stufenmodelle von Leserwerb / Leseentwicklung inkl. Vorläuferfertigkeiten	8	Identifizieren von Elementen der phonologischen Bewusstheit	6 7	Identifizieren der einfachsten Leseaufgabe auf Wortebene Ordnen von Leseaufgaben nach Lese-strategie
<i>Facette 4</i>				
Wissen über didaktische Konzepte und Hilfestellungen für den Leseunterricht	17 22	Vermittlung von Lese-strategien für das Leseverständnis Analyse von Beispielen bei Anlauttabelle	21 23 24 25	Wirksamkeit von Stilleseverfahren Hilfestellung zur Leseflüssigkeit mit Silbenbogen Maßnahmen zur Förderung Lese-strategie Wissen über effektive Förderung der Leseflüssigkeit
Σ 7 Items			Σ 8 Items	
<i>Facette 1</i>				
Wissen über Symptome von Leseschwierigkeiten	11 15	Typische Leseschwierigkeiten im Anfangsunterricht Typische Leseschwierigkeiten 3. Klasse	1 2	Analyse von Lesekompetenzen 2. Klasse Analyse von Lesekompetenzen 4. Klasse
<i>Facette 2</i>				
Wissen über Leseverläufe bei unauffällig und schwach lesenden Kindern / Wissen über typische Schwierigkeiten bei Stufenübergängen	12	Typische Symptome von Leseschwierigkeiten im Vergleich zu gut lesenden Kindern	3 4 5	Merkmale Stufe der Leseentwicklung Diagnose und Förderung LRS 2. Klasse Identifizieren einer Leseschwierigkeit auf Wortebene
<i>Facette 3</i>				
Wissen über diagnostische Verfahren und Aufgaben zur Erfassung von Kompetenzen zu Teilbereichen des Lesens	20	Diagnose von ersten lesebezogenen Auffälligkeiten	13 14 16 18	Diagnose mittels 1-Minuten-Leseprobe Analyse einer Aufgabe zur Benennungsgeschwindigkeit Spezifisches, diagnostisches Potenzial von Leseaufgaben Entwickeln einer Leseaufgabe für Kinder, die ratend lesen
Σ 4 Items			Σ 9 Items	

In Tabelle 5 werden die Items den entsprechenden Facetten zum Grundlagenwissen zugeordnet, und die in Kap. 2 ausführlich beschriebenen theoretischen Bezüge werden transparent gemacht.

Tabelle 5: Theoriebezug der Items zu Facetten des Grundlagenwissens und diagnostischen Wissens

Facetten zum Grundlagenwissen		
Item Nr.		Theoriebezug Kapitel
	<i>Facette 1</i> Wissen über linguistische Merkmale der deutschen Schriftsprache und Aspekte der Textschwierigkeit	
26	Ordnen der Satzschwierigkeit	Kap. 2.7
27	Phonemanalyse von 6 Begriffen	Kap. 2.4
28	Was ist der Lix und über welche Textmerkmale wird er berechnet	Kap. 2.7
	<i>Facette 2</i> Wissen über Prozessmodelle des Lesens und Determinanten der Lesekompetenz	
9	Zusammenhang Leseflüssigkeit und Leseverständnis	Kap. 2.3.2
10	Zusammenhang Hörverständnis Leseverständnis	Kap. 2.3.3
19	Mindestanteil an korrekt gelesenen Wörtern für unabhängiges Lesen	Kap. 2.3.2
	<i>Facette 3</i> Wissen über Stufenmodelle von Leseerwerb / Leseentwicklung inkl. Vorläuferfertigkeiten	
6	Identifizieren der einfachsten Leseaufgabe auf Wortebene	Kap. 2.5.2
7	Ordnen von Leseaufgaben nach Lesestrategie	Kap. 2.6
8	Identifizieren von Elementen der phonologischen Bewusstheit	Kap. 2.5.2
	<i>Facette 4</i> Wissen über didaktische Konzepte und Hilfestellungen für den Leseunterricht	
17	Vermittlung von Lesestrategien für das Leseverständnis	Kap. 2.3.3 / Kap.2.8.4
21	Wirksamkeit von Stilleseverfahren	Kap. 2.3.3
22	Analyse von Beispielen bei Anlauttabelle	Kap. 2.4
23	Hilfestellung zur Leseflüssigkeit mit Silbenbogen	Kap. 2.4
24	Maßnahmen zur Förderung Lesestrategie	Kap. 2.3.3
25	Wissen über effektive Förderung der Leseflüssigkeit	Kap. 2.8.3
Facetten zum diagnostischen Wissen		
Item Nr.		Theoriebezug Kapitel
	<i>Facette 1</i> Wissen über Symptome von Leseschwierigkeiten	
1	Analyse von Lesekompetenzen 2. Klasse	Kap. 2.3 / Kap.2.8
2	Analyse von Lesekompetenzen 4. Klasse	Kap. 2.3 / Kap.2.8
11	Typische Leseschwierigkeiten im Anfangsunterricht	Kap. 2.8.2
15	Typische Leseschwierigkeiten 3. Klasse	Kap. 2.8.3
	<i>Facette 2</i> Wissen über Leseverläufe bei unauffällig und schwach lesenden Kindern / Wissen über typische Schwierigkeiten bei Stufenübergängen	
3	Merkmale Stufe der Leseentwicklung	Kap. 2.6
4	Diagnose und Förderung LRS 2. Klasse	Kap. 2.8.2
5	Identifizieren einer Leseschwierigkeit auf Wortebene	Kap. 2.6
12	Typische Symptome von Leseschwierigkeiten im Vergleich zu gut lesenden Kindern	Kap. 2.8.2
	<i>Facette 3</i> Wissen über diagnostische Verfahren und Aufgaben zur Erfassung von Kompetenzen zu Teilbereichen des Lesens	
13	Diagnose mittels 1-Minuten-Leseprobe	Kap. 2.9.2
14	Analyse einer Aufgabe zur Benennungsgeschwindigkeit	Kap. 2.5.3
16	Spezifisches, diagnostisches Potenzial von Leseaufgaben	Kap. 2.9
18	Entwickeln einer Leseaufgabe für Kinder, die ratend lesen	Kap. 2.9.3
20	Diagnose von ersten lesebezogenen Auffälligkeiten	Kap. 2.8.2 /Kap. 2.9.2

5.3.2 Entwicklung der Items und Beschreibung der Itemformate

Neben der inhaltlichen Festlegung musste die Art der Befragung bestimmt werden, was eng mit dem Anspruch unterschiedlicher kognitiver Anforderungen der Items verknüpft ist. In Anlehnung an das Befragungsinstrument von Corvacho del Toro (2013) fiel die Wahl auf eine Mischung verschiedener gebundener und offener Antwortformate, die verschiedene Wissensrepräsentationen (prozedurales und deklaratives Wissen) berücksichtigen (Kap. 5.3.2). Items

müssen formalen Kriterien standhalten, und relevante Arbeitsschritte der Testkonstruktion müssen befolgt werden. Insbesondere werden bei Downing (2006) relevante Standards zur effektiven Testentwicklung beschrieben, wobei die Optimierung der Validität eine zentrale Rolle spielt. Des Weiteren sind die theoretischen und praktischen Hinweise von der Festlegung des späteren Testzwecks bis hin zur Itemkonstruktion und Itemanalyse aus zahlreichen Standardwerken in die Entwicklung des Befragungsinstruments eingegangen (Bühner, 2011; Eid & Schmidt, 2014; Moosbrugger & Kelava, 2012b; Nitko & Brookhart, 2011; Rost, 2004). Dies betraf einerseits die Entwicklung der einzelnen Items, aber auch Aspekte der Instruktion, des Layouts oder der Anordnung von Items (Moosbrugger & Kelava, 2012b; Rost, 2004).

Im Folgenden werden theoretische Grundlagen zur Test- und Itemkonstruktion erläutert, die der Entwicklung des vorliegenden Instrumentes zugrunde liegen. Es werden charakteristische Merkmale verschiedener im Instrument angewendeter Itemformate beschrieben und die Folgen auf das zu messende Konstrukt offengelegt. Dabei werden auch Überlegungen zur Kodierung und Auswertung angestellt.

Folgen des Itemformats für kognitive Anforderungen

Die Konstruktion von Items entscheidet wesentlich darüber, ob das Wissen auf einem relativ niedrigen Level abgefragt werden soll, oder ob durch die Fragetechnik oder das Itemformat höhere kognitive Kompetenzen erfragt werden sollen (Nitko & Brookhart, 2011). Die Frage nach der Abgrenzbarkeit der kognitiven Anforderungen (Fachwissen und fachdidaktisches Wissen/Können) scheint stark von der Konzeption eines Tests abhängig zu sein (Neuweg, 2014).

Wenn inhaltliches Wissen als deklaratives Wissen auf kognitiver Ebene von *Kennen/Erinnern/Abrufen* erfasst werden soll, erfordert dies Aufgaben, bei denen etwa Beispiele für Definitionen gegeben werden oder bei denen zu einem Begriff, Phänomen oder Konzept Elemente aufgezählt werden müssen und Begriffe und Konzepte erkannt werden (König, 2012). Darüber hinaus geht es beim fachdidaktischen Wissen/Können, das auf der kognitiven Ebene von *Verstehen/Analysieren/Anwenden* vorliegt, um die Verbindung dieses Wissens mit einer Problemstellung. Items, die diese Qualität von Wissen erfassen, erfordern, dass Testpersonen Konzepte und Sachverhalte charakterisieren oder erklären können. Es geht aber auch darum, einen Sachverhalt, einen Fall oder Begriffe zu vergleichen, zu kategorisieren, zu ordnen und zu interpretieren (König, 2012). In der Studie TEDS-LT wird der Begriff des *Anwendens* dem *Verstehen* zugeordnet. „Anwenden“ meint hier den Gebrauch von Prozeduren zum Üben oder zur Problemlösung (Bremerich-Vos et al., 2011, S. 54).

Offene Itemformate

Neben diesen an das Modell gebundene Überlegungen zur Wahl des Itemformats war zu entscheiden, wie sich Itemformate (unabhängig vom Inhalt) darauf auswirken, ob ein Item gut oder weniger gut gelöst wird und ob sich Items somit eignen, Wissensunterschiede zu erfassen. Intuitiv wäre davon auszugehen, dass offene Items in der Regel schwieriger zu lösen sind, da eine Reproduktion von Wissen bzw. eine selbst erzeugte Antwort verlangt wird (Moosbrugger & Kelava, 2012b). Nitko und Brookhart (2011, S. 105) zeigen indessen auf, dass sich mit diesem Format sowohl Fähigkeiten auf hohem Niveau („higher-level abilities“) als auch auf niedrigem Niveau („lower-order thinking skills“) messen lassen. So kann einerseits Wissen abgefragt werden nach dem Muster: „Wie viele Konsonanten gibt es in der deutschen Sprache?“ („knowledge of classification and categories“) oder Wissen, das nach weiteren Ausführungen verlangt, etwa nach dem Muster: „Was ist ein Graphem?“ („knowledge of terminology“). Gemäß dem Muster „knowledge of terminology“ ist das in Abbildung 13 gezeigte Item zum Grundlagenwissen konzipiert. Hier geht es um die Kenntnis des Lesbarkeitsindex, und zwar auf dem kognitiven Anforderungsniveau von *Kennen/Erinnern/Abrufen*.

Texte können sich in ihrem „Lix“ unterscheiden. Beantworten Sie untenstehende Fragen a) b). <i>Bitte nicht nur Stichwörter, sondern ganze Sätze schreiben. Falls Sie die Antwort(en) nicht wissen, schreiben Sie „ich weiss es nicht“.</i>	
a	Was bedeutet der Lix?
b	Über welche Textmerkmale wird der Lix berechnet?

Abbildung 13. Beispiel-Item 28 mit offenem Kurzantwort-Format zum Lesbarkeitsindex

Für die Beantwortung der Fragen in Abbildung 13 wird vorausgesetzt, dass Studierende den Begriff *Lix* kennen und inhaltlich umschreiben können. Beim gezeigten Item wurde davon ausgegangen, dass mit einem offenen Format neben der reinen Kenntnis des Fachbegriffs auch das Bewusstsein über die lesedidaktische Bedeutung von Oberflächen-Textmerkmalen erfasst und damit ein praktisch relevanter Aspekt abgefragt wird.

Ein Vorteil offener Antwortformate ist die Verringerung von Ratewahrscheinlichkeiten (Rost, 2004). Während bei einem Itemformat mit einer richtig-falsch-Antwortvorgabe eine Ratewahrscheinlichkeit von 50% vorliegt, ist sie bei offenen Antworten deutlich geringer oder gar nicht vorhanden (Nitko & Brookhart, 2011, S. 106). Wenn sich offene Antworten nicht eindeutig als richtig oder falsch einordnen lassen, können mehrstufige Bewertungen (z.B. falsche, halbrichtige, richtige Antwort) helfen, das Professionswissen konsistenter zu bewerten (ebd.). Ausführungen dazu sind in Kap. 5.3.4 festgehalten.

Die anfängliche Absicht, nur gebundene Formate zu verwenden, wurde u.a. aus diesen Gründen verworfen. Damit wurden die Nachteile der unökonomischeren Durchführung und Auswertung sowie der geringeren Auswertungsobjektivität offener Fragen (Moosbrugger & Kelava, 2012b) teilweise in Kauf genommen. Bereits bei der Voruntersuchung hat sich gezeigt, dass sich offene Fragen u.a. durch hohe Trennschärfen auszeichnen und dass sie sich eignen, Fähigkeiten im oberen Wissensbereich zu erfassen. Außerdem sind gebundene Formate nicht in jedem Fall geeignet, um bestimmte Antworten zu provozieren. Die Autorinnen und Autoren der FALKO-Studie (Krauss et al., 2017b) argumentieren, dass sich offene Items generell besser eignen, um die „unterrichtliche Handlungsfähigkeit“ (S. 40) und das vernetzte flexible Professionswissen abzubilden. Dies gilt insbesondere für das fachdidaktische Wissen/Können. Trotz dieses berechtigten Einwands wurden gebundene Formate aus ökonomischen Gründen und aus Gründen der höheren Auswertungsobjektivität dennoch präferiert, ungebundene Formate aber nicht a priori ausgeschlossen.

Gebundene Itemformate

Grundsätzlich gilt für alle gebundenen Itemformate, dass Antworten vorgegeben sind und dass Probandinnen und Probanden in ihren Reaktionen nicht frei sind (Moosbrugger & Kelava, 2012b). Dieser Umstand der „Gelenktheit“ führt dazu, dass der Freiraum für eigene Antworten eingeschränkt ist. Die Antwortalternativen schöpfen möglicherweise nicht alle Reaktionsmöglichkeiten aus (Bühner, 2011, S. 110; Rost, 2004, S. 61). Dieser Vorbehalt ist bei gebundenen Formaten sicher angebracht. Die mentale Repräsentation von Wissen scheint dadurch aber nicht zwingend eingeschränkt. Wie mit offenen Formaten lassen sich grundsätzlich auch mit gebundenen Formaten verschiedene Levels kognitiver Verarbeitung von Wissen und Können erfassen.

Bei gebundenen Formaten muss das Problem der Plausibilität von Distraktoren gelöst werden, d.h., dass falsche aber plausible Antwortmöglichkeiten von der richtigen Antwort ablenken sollen (Bühner, 2011). Um zu gewährleisten, dass Distraktoren als *echte* Antwortalternativen wahrgenommen werden, wurden bei der Formulierung die Grundsätze nach Moosbrugger & Kelava, (2012b) und Nitko und Brookhart (2011) berücksichtigt. Inhaltlich basieren plausible Distraktoren dabei auf häufig vorkommenden Fehlern von Studierenden, d.h. auf verbreitetem Fehlwissen. Je ähnlicher sich vorgegebene Antwortkategorien in den gebundenen Formaten sind, desto schwieriger ist der Auswahlprozess für die befragten Personen (Rost, 2004). Die Schwierigkeit eines Items hängt zudem von der Anzahl plausibler Distraktoren ab, d.h. von der Wahrscheinlichkeit, auch zufällig zur richtigen Lösung zu kommen (Bühner, 2011). Nitko &

Brookhart, (2011) schlagen vor drei bis fünf funktionierende und plausible Antwortalternativen zu formulieren. Sowohl unplausible Distraktoren als auch plausible Antworten führen zu einer unerwünschten Erhöhung der Ratewahrscheinlichkeit, weshalb die Identifikation dieser Distraktoren bzw. Antworten entscheidend ist (ebd.). In der Vorstudie erfolgte die Beurteilung der Plausibilität im Rahmen der qualitativen Validierung (Kap. 5.3.3). Durch das laute Kommentieren von Aufgaben im kognitiven Pretestingverfahren konnten unplausible Antwortalternativen identifiziert werden (Bühner, 2011).

Die Ratewahrscheinlichkeit lässt sich zwar durch verschiedene Strategien reduzieren, aber nicht ganz vermeiden (Bühner, 2011; Moosbrugger & Kelava, 2012b). Um die Ratewahrscheinlichkeit zu reduzieren, wurden im vorliegenden Befragungsinstrument in der Regel neben der korrekten Antwort mindestens vier falsche Antwortmöglichkeiten vorgegeben. Die Ratewahrscheinlichkeit liegt dementsprechend bei 20% (Bühner, 2011, S. 120). Die Ratewahrscheinlichkeit könnte durch die Erweiterung von Antwortkategorien weiter gesenkt werden. Auch Reliabilität und Validität würden sich dank dieser Maßnahme verbessern. Allerdings ist die Steigerung der Reliabilität und Validität ab sieben Kategorien nur gering (ebd.), so dass die Begrenzung auf fünf Antwortalternativen auch unter dem Blickwinkel der Ökonomie gerechtfertigt erscheint.

Trotz kritischer Einwände gegenüber gebundenen Formaten haben diese einige Vorteile vorzuweisen, die deren Einsatz rechtfertigen (Nitko & Brookhart, 2011). Zu den wesentlichen Vorteilen gehören die Durchführungs- und Auswertungsobjektivität sowie die weitaus größere Auswertungsökonomie gegenüber offenen Fragen (Rost, 2004).

Verwendete Itemformate im Befragungsinstrument

Neben dem ungebundenen offenen Kurzantwort-Format wurden fünf verschiedene Typen gebundener Itemformate gewählt. Die folgende Tabelle 6 zeigt eine Übersicht aller in der Hauptstudie verwendeten Formate. Im Anschluss daran werden die geschlossenen Formen kurz beschrieben. Die Zuordnung der Formate zu den Feldern des Modells des lesebezogenen Wissens/Könnens soll die Unabhängigkeit von Format und Art der kognitiven Anforderung (mentale Repräsentation von Wissen/Können) aufzeigen.

Tabelle 6: Verteilung der Itemformate im Modell in der Hauptstudie

Inhaltsbereich und Format	Kognitive Anforderung			
	Item Nr.	Fachwissen Kennen/Erinnern/Abrufen	Item Nr.	Fachdidaktisches Wissen/Können Verstehen/Analysieren/Anwenden
Grundlagenwissen				
Offene Kurzantwort	28	Was ist der <i>Lix</i> und über welche Textmerkmale wird er berechnet	23	Hilfestellung zur Leseflüssigkeit mit Silbenbogen
Single-Choice	9	Zusammenhang Leseflüssigkeit und Leseverständnis	6	Identifizieren der einfachsten Leseaufgabe auf Wortebene
	10	Zusammenhang Hörverständnis Leseverständnis	7	Ordnen von Leseaufgaben nach Lesestrategie
	17	Vermittlung von Lesestrategien für das Leseverständnis	25	Wissen über effektive Förderung der Leseflüssigkeit
	19	Mindestanteil an korrekt gelesenen Wörtern für unabhängiges Lesen		
Multiple-Choice	8	Identifizieren von Elementen der phonologischen Bewusstheit	24	Maßnahmen zur Förderung der Lesestrategie
Antwortbatterien mit richtig-falsch Antwort	22	Analyse von Beispielen bei Anlauttabelle	21	Wirksamkeit von Stilleseverfahren
			27	Phonemanalyse von 6 Begriffen
Problembasierte Audio-Items	–		–	
Zu- und Umordnungsaufgaben	–		26	Ordnen der Satzschwierigkeit
Diagnostisches Wissen				
Offene Kurzantwort	–		13	Diagnose mittels 1-Minuten-Leseprobe
			14	Analyse einer Aufgabe zur Benennungsgeschwindigkeit
			18	Entwickeln einer Leseaufgabe für Kinder, die ratend lesen
Single-Choice	11	Typische Leseschwierigkeiten im Anfangsunterricht	5	Identifizieren einer Leseschwierigkeit auf Wortebene
	12	Typische Leseschwierigkeiten im Vergleich zu gut lesenden Kindern		
	15	Typische Symptome von Leseschwierigkeiten 3. Klasse		
	20	Diagnose von ersten lesebezogenen Auffälligkeiten		
Multiple-Choice	–		–	
Antwortbatterien mit richtig-falsch Antwort	–		–	
Problembasierte Audio-Items	–		1	Analyse von Lesekompetenzen 2. Klasse
			2	Analyse von Lesekompetenzen 4. Klasse
			3	Merkmale Stufe der Leseentwicklung
			4	Diagnose und Förderung LRS 2. Klasse
Zu- und Umordnungsaufgaben	–		16	Spezifisches, diagnostisches Potenzial von Leseaufgaben

Single-Choice: Bei diesem Format kann bei mehr als zwei Antwortalternativen nur *eine* Antwort die richtige sein. Das Format wurde im Instrument weitaus am häufigsten verwendet. Eine Variation von Single-Choice- Aufgaben sind „Best-Answer-Items“ (Nitko & Brookhart, 2011, S. 140). Diese Variante wurde im Instrument insgesamt viermal (Item 15, 16, 17, 25) gewählt. Bei Best-Answer-Items ist jede vorgeschlagene Antwort in gewisser Weise korrekt, aber nur

eine Antwort kann die beste sein. Das Format hat den Vorteil, dass eine differenzierte Entscheidung getroffen werden muss, womit vertieftes Wissen abgefragt werden kann (ebd.). Items dieses Formats setzen voraus, dass Studierende ein gewisses Niveau an Wissen und Können erreicht haben, denn solche Items sind für weniger kompetente Studierende unlösbar.

Multiple-Choice: Bei diesem Format besteht die Möglichkeit, alle Antworten, eine Auswahl oder keine Antwort zu wählen. Die Anzahl der möglichen richtigen Antworten kann dabei angegeben oder offengelassen werden. Im letzteren Fall sinkt die Ratewahrscheinlichkeit bei fünf Antwortkategorien auf 1/32 (Rost, 2004, S. 63). Im vorliegenden Instrument kommt dieses Format lediglich bei Item 8 vor. Bei allen anderen solchen Items wurden zwei richtige Antworten von fünf erwartet, was zu einer Ratewahrscheinlichkeit von 10% führt (Bühner, 2011, S. 120; Rost, 2004, S. 63). Für die Kodierung muss berücksichtigt werden, dass hier in der Regel die richtige Kategorienkombination als Itemlösung kodiert wird (Rost, 2004).

Antwortbatterien mit Richtig-Falsch-Antwort: Richtig-Falsch-Aufgaben bestehen aus lediglich zwei Antwortmöglichkeiten im Sinne von Ja-Nein-Fragen (Bühner, 2011). Die Ratewahrscheinlichkeit liegt hier bei 50%, was insbesondere bei Leistungstests ein Problem darstellt (ebd., S. 117). Um das Problem der hohen Ratewahrscheinlichkeit zu entschärfen, wurde deshalb ein Set von vier Fragen vorgelegt, wobei jede Frage einzeln mit „stimmt“ oder „stimmt nicht“ beantwortet werden musste. Die Wahrscheinlichkeit, die richtige Kombination zu erraten, liegt bei rund 8%. Diesem Format wird auch Item 27 zugeordnet, bei dem bei sechs Wörtern jeweils die richtige Anzahl Phoneme angekreuzt werden muss.

Zu- und Umordnungsaufgaben: Korrekterweise sind dies zwei unterschiedliche Formate. Im Fall von Zuordnungsaufgaben geht es darum, einer Liste von Aussagen eine Reihe von Antworten zuzuordnen (Nitko & Brookhart, 2011). Die Anzahl an Aussagen und zuzuordnenden Aussagen kann identisch sein (perfektes matching). Es können aber auch mehr Aussagen als Antworten vorgegeben sein oder umgekehrt (ebd.). Die zufällige Beantwortung ist hier unproblematisch, sofern die Anzahl der zuzuordnenden Antwortalternativen größer ist als die der Prämissen (Moosbrugger & Kelava, 2012b). Da sich in der Voruntersuchung die Anweisung bei nicht perfektem matching als verwirrend herausgestellt hat, wurde ein Item für die Hauptstudie vereinfacht, so dass die Anzahl Aussagen und zugehörigen Antworten übereinstimmen. Das Format hat den Vorteil, dass Kenntnisse und Wissen platzsparend, ökonomisch und objektiv überprüft werden können (ebd.).

Bei *Umordnungsaufgaben* geht es darum, vorgegebene Fragmente (Worte, Sätze, Zahlen usw.) in eine richtige Reihenfolge zu bringen (Bühner, 2011; Moosbrugger & Kelava, 2012b).

Problembasierte Audio-Items: Dieses Frageformat basiert auf authentischen Audiobeispielen lesender Kinder, die den befragten Personen als Aufgabenstimulus abgespielt werden. Im Anschluss daran mussten die Studierenden in der Voruntersuchung beurteilen, welche lesebezogenen Kompetenzen die Kinder mehrheitlich zeigen und wie sie die Lesekompetenz der Kinder gesamthaft einschätzen. In einem Fall mussten zudem zwei didaktische Maßnahmen gewählt werden, die bei der Überwindung des lesebezogenen Problems helfen könnten. Die Antwortalternativen wurden im gebundenen Multiple-Choice-Format vorgelegt.

Das Format nimmt Bezug auf die Idee von Fallvignetten, wie sie in anderen Studien zur Erfassung didaktischer Kompetenzen von Lehrpersonen verwendet wurden (siehe Studien: Bruckmaier et al., 2013; IteL, 2015; Oser, Heinzer & Salzmann, 2010; Rutsch, 2016). Allen Studien liegt die Annahme zugrunde, dass die Komplexität von Unterrichtssituationen durch alltagsnahe realistische Beispiele wiedergegeben werden kann (IteL, 2015) und dass bei den befragten Personen unterrichtsnahe kontextgebundene Reaktionen provoziert werden (Lindmeier, 2013). Fallvignetten eignen sich besonders, um das praktische Wissen abzufragen, das „üblicherweise an Fälle, Episoden und Skripts gebunden ist“ (Baumert & Kunter, 2011a, S. 35). Bei diesem Format wird eine höhere Denkleistung („higher-order thinking“) verlangt, da Aufgaben in einem Kontext zu lösen sind (Nitko & Brookhart, 2011, S. 178).

5.3.3 Qualitative Validierung des Befragungsinstruments

Zur inhaltlichen und formalen Optimierung der Items wurden zwei kommunikative Validierungsverfahren – eine Expertinnen- und Expertenvalidierung und ein kognitives Pretestigungsverfahren – durchgeführt und ungünstige Items in einem mehrstufigen Prozess sukzessive reduziert. Da beiden Verfahren *vor* dem Einsatz des Befragungsinstrumentes durchgeführt wurden, konnten Items aufgrund bestimmter Kriterien aus dem Itempool entfernt werden.

5.3.3.1 Befragung von Expertinnen und Experten

Die Form der kommunikativen Validierung durch Expertinnen und Experten ist in zahlreichen Studien Bestandteil der Testentwicklung (TEDS-M: Blömeke & König, 2010; COACTIV-Studie: Krauss et al., 2011; FALKO-D: Pissarek & Schilcher, 2017; Paderborner Kompetenztest: Vogelsang & Reinhold, 2013). Ziel dieser Entwicklungsphase war es, Items hinsichtlich ihrer praktischen und fachlichen Relevanz, Eindeutigkeit und Verständlichkeit zu optimieren. Daneben wurde das Expertinnen- und Expertenwissen zur Entwicklung eines prototypischen Kodiermanuals genutzt. Dabei wurden keine quantitativen cut-off-Werte festgelegt, sondern Items wurden unter theoretischen und normativen Gesichtspunkten ausgewählt (Pissarek & Schilcher, 2017).

Die Gruppe der Expertinnen und Experten setzte sich in der Voruntersuchung aus vier Fachdidaktikerinnen bzw. Fachdidaktikern zusammen, die sich beruflich und wissenschaftlich mit dem Themenbereich der Lesedidaktik beschäftigen (Rutsch, 2016). Zusätzlich wurden die Items hinsichtlich Kodierbarkeit und formaler Kriterien im Rahmen eines Doktoranden-Kolloquiums diskutiert. Für die Hauptstudie wurden fünf neue Expertinnen und Experten befragt, die profunde Kenntnisse im Bereich des Anfangsleseunterrichts hatten.

In Anlehnung an das Vorgehen bei FALKO-D (Pissarek & Schilcher, 2017, S. 87) wurden die Items hinsichtlich folgender Kriterien beurteilt:

- Welche Items eignen sich besonders gut, um das Wissen von Lehrpersonen im Bereich des Leseanfangsunterrichts zu erfassen? (Augenscheinvalidität)
- Sind die Inhalte der Items zentral für den Leseunterricht der ersten bis vierten Klasse? (Dabei wurde auch nach dem Kriterium der praktischen und/oder theoretischen Relevanz gefragt)
- Sind die Lösungen eindeutig (Items in geschlossenem Format), und entsprechen gute Lösungen (Items im offenen Format) dem Erwartungshorizont der Expertinnen und Experten?

Weiter musste durch die Expertinnen und Experten abgesichert werden, dass die Items dem aktuellen Stand der Theorie entsprachen. Die Expertinnen und Experten sollten bei der Auswahl der Items berücksichtigen, ob abgefragte Themen in Lehrbüchern, Vorlesungen und in der Forschung zentral behandelt werden. Items zu Themen, die auch in der Theorie eher selten oder sehr kontrovers diskutiert werden, wurden aus dem Itempool ausgeschlossen.

Bei gebundenen Antwortformaten musste die richtige Lösung durch die Expertinnen und Experten eindeutig identifiziert werden können. War dies nicht der Fall, wurden entsprechende Items vor der Datenerhebung überarbeitet oder ggf. aus dem Itempool ausgeschlossen. Bei offenen Fragen wurde ein theoretischer Erwartungshorizont über angemessene richtige Antworten formuliert, die zudem durch die Literatur abgesichert waren (Krauss et al., 2011).

5.3.3.2 Kognitives Pretesting

Im Rahmen der qualitativen Validierung des Instruments wurde zu Fragen der Verständlichkeit und Eindeutigkeit eine kognitive Pretest-Technik eingesetzt (Blömeke & König, 2010). Kognitive Pretest-Techniken werden zum Einsatz gebracht, um Einblicke in die kognitiven Prozesse zu erhalten, die zur Beantwortung einer Frage führen (Prost, 2014). Eine kognitive Pretest-Technik ist das „think-aloud“-Verfahren, bei dem potenzielle Studienteilnehmende aufgefordert werden, während dem Beantworten der Fragen „laut zu denken“ und alle Überlegungen aufzuführen, die zur Beantwortung der Frage führen (Moosbrugger & Kelava, 2012b; Prost,

2014). Ziel dieser protokollierten und ausgewerteten Rückmeldungen war es, Probleme bei der Formulierung von Items und der Handhabung zu beheben (Antonietti, Moreau, Moser & Ramseier, 2006). Kriterien für die Auswahl geeigneter Items waren Aspekte der Verständlichkeit von Items sowie die Einschätzung, ob Antworten bei offenen Fragen in die intendierte Richtung gingen und ob Distraktoren plausibel erschienen. Die potenziellen Teilnehmenden der Vorstudie mussten die spätere Zielstichprobe repräsentieren und gleichmäßig auf die Substichproben verteilt sein (Prost, 2014). In der *Hauptstudie* wurden sechs potenzielle Probandinnen und Probanden befragt.

5.3.4 Kodierung der Items und Entwicklung des Kodiermanuals

Neben der Entwicklung der Items wurde zeitgleich die „kriterienorientierte Interpretationsgrundlage“ geschaffen, d.h. das Kodiermanual erstellt (Jandl, 2016, S. 144). Die Entwicklung erfolgte einerseits auf Grundlage der theoretischen Bezüge der einzelnen Items. Andererseits wurden aber auch Rückmeldungen aus der Validierung durch Expertinnen und Experten berücksichtigt. Das Vorgehen folgt damit dem Vorgehen anderer Studien (z.B. FALKO-D Pissarek & Schilcher, 2017).

Die Punktevergabe erfolgte bei den meisten Items in zwei Schritten: Bei Items mit gebundenem Format wurde im ersten Schritt für jedes korrekt gesetzte Kreuz und jedes korrekt *nicht* gesetzte Kreuz ein Punkt vergeben. Bei vier Distraktoren und einer richtigen Antwort konnten z.B. fünf Punkte erreicht werden. Diese Punkte wurden anschließend in ein zweistufiges oder dreistufiges Scoring überführt. Das Scoring der Antworten erfolgte bei 14 Items zweistufig (dichotom), d.h. mit null Punkten für die *falsche Antwort* und einem Punkt für die *richtige Antwort* (Abbildung 14).

Ein Kind am Ende der 2. Klasse liest in zügigem Tempo < r'o's'a' > (jeden Buchstaben einzeln). Womit hat dieses Kind Mühe? (Eine Antwort ankreuzen)		
Das Kind hat Mühe...		
a	mit der auditiven Unterscheidung der Laute	<input type="checkbox"/>
b	mit dem Lautieren der Buchstaben	<input type="checkbox"/>
c	mit dem Zusammenlauten von Buchstaben	<input type="checkbox"/>
d	mit dem auditiven Arbeitsgedächtnis	<input type="checkbox"/>
e	mit der phonologischen Bewusstheit	<input type="checkbox"/>

Kodierung:
0: 1-4 falsche Antworten
1: 5 korrekte Antworten

Abbildung 14. Beispiel-Item 5, Single-Choice mit einer korrekten Antwortmöglichkeit (max. 5 Punkte)

Dieses Prinzip kam bei allen Single-Choice-Aufgaben mit nur *einer* richtigen Antwort und bei zwei offenen Items zum Tragen. Bei 14 Items wurde ein dreistufiges Scoring nach dem Partial-Credit-Modell vorgenommen. Dabei wurden auch teilrichtige Lösungen berücksichtigt und entsprechend null Punkte für die *falsche Lösung*, ein Punkt für *teilrichtige Lösungen* und zwei

Punkte für *vollständig richtige Lösungen* vergeben (Nitko & Brookhart, 2011). Das Partial-Credit-Modell kam überall dort zum Einsatz, wo teilrichtige Lösungen möglich waren, d.h. wo *mehrere* Antworten verlangt wurden (Multiple-Choice-Aufgaben) und wo neben der richtigen Lösung auch *eine* falsche Antwort angekreuzt wurde. Der zweite Punkt steht für ein höheres Wissen. Bei Zu- und Umordnungsaufgaben wurde eine Verwechslung noch als halbrichtig gewertet. Bei Items, bei denen mehrere Antworten angekreuzt werden mussten, galt konsequent eine 1-Fehler-Toleranz, um zumindest einen Punkt von maximal zwei Punkten zu erreichen. Anhand eines Beispiels wird im Folgenden ein Einblick in die Kodierarbeit für Items mit Partial-Credit-Kodierung gegeben (Abbildung 15).

<p>Das Kind steht am Anfang 1. Klasse und verwendet die weitgehend entfaltete alphabetische Lesestrategie. An welchen der folgenden Merkmale hören Sie in diesem Beispiel, dass das Kind diese Strategie dominant verwendet? (Mehrere Antworten ankreuzen)</p> <p>Alter: 6 Jahre 5 Monate Stufe: Anfang 1. Klasse (September)</p>			<p><i>Kodierung:</i> 0: 0-4 korrekte Antworten 1: 5 korrekte Antworten 2: 6 korrekte Antworten</p>
a	Das Kind liest Wörter teilweise gedehnt.	<input type="checkbox"/>	
b	Das Kind liest kurze Wörter zügig und fehlerlos.	<input type="checkbox"/>	
c	Dem Kind gelingt es während dem Lesen mehrheitlich, größere Worteinheiten (z.B. Silben) zu erfassen.	<input type="checkbox"/>	
d	Das Kind zeigt eine verzögerte Sinnentnahme.	<input type="checkbox"/>	
e	Das Kind kann Wörter mehrheitlich buchstabenweise vollständig erlesen.	<input type="checkbox"/>	
f	Das Kind liest Wörter mehrheitlich als Ganzes (Sichtwortschatz).	<input type="checkbox"/>	

Abbildung 15. Beispiel-Item 3, Audio-Item mit mehreren Antwortmöglichkeiten (max. 6 Punkte)

Da das Befragungsinstrument nach Prinzipien von Niveautests konzipiert ist, für dessen Bearbeitung ausreichend Zeit zur Verfügung steht, wurden nicht bearbeitete Items als falsch (0 Punkte) gewertet (Bochnik, 2017).

Zur Kodierung von Items mit offenem Format wurde nach dem Vorbild vorausgehender Studien (Krauss et al., 2011; Pissarek & Schilcher, 2017; Rutsch, 2016) auf Grundlage der theoretischen Herleitungen der Items und der Expertinnen- und Expertenbefragung ein provisorisches Lösungsmuster festgelegt. Für jedes Item wurde ein theoretischer Erwartungshorizont über mögliche zu erwartende richtige oder falsche Antworten formuliert (Krauss et al., 2011). Aufgrund der Rückmeldungen durch die Expertinnen und Experten zu den erwarteten Antworten wurde das Kodiermanual erweitert und differenziert. Im nächsten Schritt wurden die ersten „echten“ Antworten analysiert und das Kodierschema mit den erhaltenen Antworten sukzessive angepasst (ebd.). Bei zwei offenen Items wurde zwischen verschiedenen Qualitäten von Antworten unterschieden und eine genaue Kodieranleitung erstellt, die diese qualitativen Unterschiede abbildet und anschließend eine Partial-Credit-Kodierung erlaubt (Abbildung 16).

Texte können sich in ihrem "Lix" unterscheiden. Beantworten Sie untenstehende Fragen a) b). <i>Bitte nicht nur Stichwörter, sondern ganze Sätze schreiben. Falls Sie die Antwort(en) nicht wissen, schreiben Sie „ich weiss es nicht“.</i>	
a	Was bedeutet der Lix?
b	Über welche Textmerkmale wird der Lix berechnet?

Abbildung 16. Beispiel-Item 28 mit offenem Antwortformat

Für die Teilaufgaben a) und b) wurde jeweils maximal ein Punkt vergeben, was bei vollständig korrekter Lösung zwei Punkte ergibt.

Kodierung Teilaufgabe a)

- 0: Keine Antwort oder Antwort, die von untenstehender Antwort abweicht
 1: Lesbarkeitsindex
 Leseindex
 Index für die Textschwierigkeit
 Leseschwierigkeitsindex
 Schwierigkeit des Textes

Kodierung Teilaufgabe b)

- 0: Keine Antwort oder Antwort, die von untenstehender Antwort abweicht bzw. unvollständige Antwort
 Schwierigkeit [ohne weitere Ergänzung]
 1: Satzlänge und Wortlänge [beides muss vorkommen!]
 (Durchschnittliche) Satzlänge und Anzahl lange Wörter
 (Durchschnittliche) Satzlänge und prozentualer Anteil an Wörtern mit mehr als 6 Buchstaben
 Schwierigkeit [und Präzisierung im Sinne] Wörter mit mehr als 6 Buchstaben [wenn nur Schwierigkeit, dann falsch]
 Anzahl Wörter in einem Satz und Wortlänge [oder Wortlänge umschrieben im Sinne] Anzahl Buchstaben in einem Wort.

Qualität der Kodierung bei offenem Itemformat

Um eine hohe Interraterreliabilität zu erreichen, erforderte die dreistufige Kodierung bei offenen Items eine exakte Kodieranweisung und gründliche Schulung der Raterinnen und Rater (Böhme, Bremerich-Vos & Robitzsch, 2009).

In der Voruntersuchung wurden für diesen Prozess drei Items mit offenem Antwortformat von zwei Raterinnen und Ratern unabhängig voneinander kodiert. In der Hauptstudie waren es drei Rater bei vorliegenden fünf Items. Für die Berechnung der Übereinstimmungsreliabilität wurden in der Voruntersuchung und Hauptstudie gut 20% des gesamten Materials von allen Raterinnen und Ratern kodiert. Dies entspricht dem in der Fachliteratur empfohlenen Anteil (Lombard, Snyder-Duch & Bracken, 2002; Mayring, 2003). Um ein repräsentatives Datenset zu erhalten, wurden die Fragebögen zufällig gezogen, unter der Bedingung, dass Antworten aller Substichproben anteilmäßig gleich vertreten waren. Für die Übereinstimmung wurde das Maß Cohens Kappa (κ) berechnet, das auch zufällige Übereinstimmungen zwischen Kodierenden berücksichtigt (Böhme et al., 2009; Bortz & Lienert, 2003). Neben Cohens-Kappa gilt Krippendorff's Alpha (r_a) als geeignet, wenn mehrere Rater dasselbe Datenmaterial kodieren

(Lombard et al., 2002). Das Maß Cohens Kappa (κ) gilt wegen seiner Genauigkeit und der Überwindung verschiedener Nachteile anderer Übereinstimmungskoeffizienten (Hayes & Krippendorff, 2007) als einschlägige Messgröße. Für den korrigierten Kappa-Koeffizienten (paarweiser Vergleich) wird dabei der Schwellenwert von $\kappa > .75$ angenommen (Bortz & Döring, 2002, S. 277; Wirtz & Caspar, 2002, S. 57). Für Krippendorff's Alpha wird meist ein Wert von $r_\alpha > .80$ als Maß akzeptabler Übereinstimmung angegeben (Lombard et al., 2002, S. 593). Der Wert ist allerdings ausgesprochen konservativ, und Korrekturen auf Werte von $r_\alpha > .75$ sind bei Kodes mit hoher Interpretationsleistung gerechtfertigt (Müller & Buchs, 2014). Da die Übereinstimmungswerte in der Hauptstudie anfänglich bei zwei Items in einem unbefriedigenden Bereich lagen, wurden die abweichenden Kodierungen diskutiert, und es wurde versucht, ein Konsens herzustellen. Kategorienbeschreibungen musste daraufhin erweitert und präzisiert werden (Krauss et al., 2011). Zur inhaltlichen Absicherung wurde das überarbeitete Kodierschema nochmals mit drei Expertinnen diskutiert, und die Übereinstimmung wurde an einer weiteren Fragebogen-Stichprobe neu berechnet. Die folgende Tabelle 7 zeigt, dass die Übereinstimmungsmasse bei allen Items in einem sehr guten ($\kappa > .75$) bis guten ($\kappa = .60 - .75$) Wertebereich liegen (Döring & Bortz, 2016). Dieselbe Interpretation gilt auch für Krippendorff's Alpha. Damit ist die Grundlage für eine möglichst objektive Auswertung der offenen Items geschaffen.

Tabelle 7: Übereinstimmungsreliabilität bei Items mit offenem Format in der Hauptuntersuchung

Item Nr.	Kappa-Koeffizient	Krippendorff's Alpha (nominal)
13	.72	.72
14	.84	.84
18	.78	.77
23	.95	.95
28	.89	.89

Anmerkung. Alle Übereinstimmungen wurden für die drei Raterinnen bzw. Rater paarweise berechnet und daraus der Mittelwert bestimmt (average pairwise)

5.4 Durchführung der Datenerhebung und Beschreibung der Stichprobe

5.4.1 Durchführung der Datenerhebung in der Vorstudie und Hauptstudie

Rekrutierung der Studienteilnehmenden und allgemeine Anforderung an die Stichprobe

Die *Vorstudie* wurde an drei Pädagogischen Hochschulen und Ausbildungsinstitutionen für Schulische Heilpädagogik der Schweiz durchgeführt. Es wurden Daten von 181 Studierenden und 22 Laien erfasst, denen 24 Items vorgelegt wurden. In der Voruntersuchung stammten die Laien nicht aus dem schulischen Bereich, sondern aus dem privaten Umfeld und waren dementsprechend keine Studierenden.

In die *Hauptstudie* wurden zwei weitere Ausbildungsinstitutionen für Schulische Heilpädagogik einbezogen. Insgesamt wurden die Daten von 514 Studierenden erfasst, die 28 Items vorgelegt bekamen. Sieben Personen wurden von den Analysen ausgeschlossen, da sie mindestens ein Drittel der Fragen gar nicht oder nicht gemäß Anweisung beantwortet hatten. Damit reduzierte sich der Datensatz auf 507 auswertbare Fragebögen. In der Hauptstichprobe stammten die Laien aus dem schulischen Umfeld und bestanden aus einer Gruppe von Studierenden, die am Anfang der Grundausbildung zur Primarlehrperson im Bachelorstudiengang standen. Befragt wurden Studierende, die einerseits die Grundausbildung zur Lehrperson auf Primarstufe absolvierten (Bachelorstudiengang) und Studierende, die einen Studiengang in Schulischer Heilpädagogik (Masterstudiengang) besuchten.

Die Anforderungen an eine Stichprobe richten sich nach dem späteren Verwendungszweck eines Instrumentes. Im vorliegenden Fall wurde beabsichtigt, ein Befragungsinstrument zu entwickeln, das ein möglichst breites Spektrum an Wissen von Lehrpersonen mit unterschiedlicher Berufsbiografie erfassen sollte. Entsprechend dem Anwendungszweck bei dieser Zielgruppe wurde bei der Auswahl der Stichprobe auf eine hohe Heterogenität geachtet.

Eine weitere Grundlage der Stichprobenziehung war der Expertiseansatz, der davon ausgeht, dass sich das Professionswissen theoretisch vom Laienstatus bis zum Expertenstatus weiterentwickelt (Kap. 3.1.2). Grundannahme war, dass sich Wissensunterschiede, die durch die Berufserfahrung und durch den Ausbildungshintergrund verursacht werden, sensitiv erfassen lassen. Aus diesem Grund musste eine genaue Zuordnung der Studienteilnehmenden in Substichproben nach relevanten berufsbiografischen Kriterien erfolgen. In der vorliegenden Untersuchung wurde neben der Berufserfahrung und besuchten Weiterbildungen auch erfasst, ob die Studienteilnehmenden eine Ausbildung in Schulischer Heilpädagogik absolviert hatten. Im Vergleich zu Studierenden, die lediglich eine Grundausbildung zur Lehrperson absolviert haben, müssten diese Studierenden über ein ausgereifteres theoretisches Wissen zu Grundlagen des Lesens und

Diagnostik bei leseschwachen Kindern verfügen. Entsprechend erfolgte eine Klassifizierung der Studienteilnehmenden in drei Substichproben:

- a) *Studierende ohne Berufserfahrung*, die am Ende ihrer Ausbildung zur Primarlehrperson (Primar_{ohneBE}) standen und alle im Studium angebotenen Kurse im Bereich Deutschdidaktik und Umgang mit Lese-Rechtschreibschwierigkeiten (LRS) besucht haben sollten.
- b) *Studierenden, die ganz am Anfang des Studiums in Schulischer Heilpädagogik* standen und gemäß Zulassungsbedingungen mindestens zwei Jahre Berufserfahrung hatten. Diese Gruppe wurde als Lehrpersonen mit Berufserfahrung (Primar_{mitBE}) aber noch ohne Zusatzwissen im Bereich Schriftspracherwerb und Umgang mit LRS klassifiziert, da sie am Anfang des Studiums noch keinerlei Zusatzwissen mitbrachten.
- c) *Studierende der Schulischen Heilpädagogik am Ende der Ausbildung*, die alle Kurse in Deutschdidaktik und im Umgang mit LRS besucht hatten und gemäß den Zulassungsbedingungen Berufserfahrung als Lehrperson mitbrachten.

Eine Laiengruppe bildete zudem die Kontraststichprobe, die nicht in die Skalenbildung einbezogen wurden, sondern lediglich als Kontrollgruppe diente. Eine genaue Stichprobenbeschreibung folgt im nächsten Kapitel.

Durchführung

Der formale Ablauf der Erhebung erfolgte in der Vor- wie in der Hauptuntersuchung identisch. Die Studiengangleitenden der jeweiligen Ausbildungsinstitute wurden schriftlich angefragt, ob sie den Kontakt zu Dozierenden herstellen könnten, die Studierende unterrichteten, die für die Datenerhebung in Frage kommen könnten. Eine genauere Unterscheidung der drei Substichproben erfolgte post-hoc über Kriterien, die mit einem Kurzfragebogen im Rahmen der Datenerhebung erfasst wurden.

Die Studierenden wurden im Vorfeld über den Untersuchungszweck und über die Freiwilligkeit zur Teilnahme informiert. Die Testleiterin – sowie in einem Fall eine geschulte Forschungsassistentin – führte die Erhebung jeweils im Rahmen einer obligatorischen Lehrveranstaltung mit allen Studierenden als Gruppenverfahren durch. Der Fragebogen lag in einem klassischen Paper-Pencil-Format vor, und die Erhebungen erfolgten nach einem standardisierten Ablauf. Die vier Audio-Items wurden nach standardisierten Vorgaben jeweils zu Beginn der Erhebung zeitgleich von allen gelöst. Anschließend beschäftigten sich die Studierenden mit den Items in ihrem eigenen Tempo. Für die Beantwortung der Fragen standen 60 Minuten zur Verfügung. Trotz dieser zeitlichen Beschränkung kann das Befragungsinstrument als Power-Test gelten (Bühner, 2011; Krauss et al., 2011), da die Zeit zur Bearbeitung großzügig bemessen war, so

dass die Studienteilnehmenden innerhalb der vorgesehenen Zeit fertig wurden. Die Dozierenden sowie die Studierenden erhielten eine schriftliche Rückmeldung zu den erzielten zusammengefassten und anonymisierten Ergebnissen der eigenen Kohorte.

5.4.2 Stichprobe

Beschreibung der Stichprobe in der Hauptstudie

Die Gesamtstichprobe umfasste nach Ausschluss der sieben oben genannten Studierenden 507 Studienteilnehmende. Für die Skalierung des Instrumentes wurden die Laien nicht einbezogen. Die Daten dieser Gruppe (Kontraststichprobe) wurden lediglich zu Validierungszwecken verwendet. Für die Entwicklung des Instrumentes wurden somit nur die Daten von 443 Personen berücksichtigt. Die Stichprobe wurde aufgrund der in einem Kurzfragebogen erfassten personenbezogenen Daten zunächst in fünf Substichproben und eine Kontraststichprobe eingeteilt (Tabelle 8).

Tabelle 8: Inhaltliche Beschreibung der Substichproben nach Kriterien der Berufsbiografie

K	Substichprobe	Beschreibung
I	Primarlehrpersonen ohne Berufserfahrung am Ende des Studiums (Primar _{ohneBE}) <i>n</i> = 190	Studierende stehen am Ende des Bachelorstudiengangs zur Grundschullehrperson (5. Semester). Das Grundlagen-Fachwissen zum Schriftspracherwerb und Umgang mit LRS ^a wurde im Rahmen der Ausbildung erworben. Studierende haben noch keine Berufserfahrung haben aber im Rahmen der Ausbildung ev. bereits Praktika absolviert.
II	Primarlehrpersonen mit Berufserfahrung (Primar _{mitBE}) <i>n</i> = 127	Studierende stehen am Beginn der Ausbildung im Masterstudiengang in Schulischer Heilpädagogik (1. Semester). Das Grundlagen-Fachwissen zum Schriftspracherwerb und Umgang mit LRS wurde im Rahmen der Ausbildung erworben. Es wurde noch kein Zusatzwissen erworben. Studierende verfügen über mind. 1-2 Jahre Berufserfahrung im Fach Deutsch auf Primarstufe.
III	SHP ^b mit Berufserfahrung (SHP) <i>n</i> = 105	Studierende stehen am Ende der Ausbildung des Masterstudiengangs in Schulischer Heilpädagogik (5. Semester). Ein vertieftes Fachwissen zum Schriftspracherwerb, Umgang mit LRS wurde im Rahmen der Zusatzausbildung erworben. Studierende verfügen über mind. 1-2 Jahre Berufserfahrung im Fach Deutsch auf Primarstufe.
IV	SHP ohne Berufserfahrung im Fach Deutsch auf Primarstufe <i>n</i> = 9	Studierende stehen am Ende der Ausbildung des Masterstudiengangs in Schulischer Heilpädagogik (5. Semester). Ein vertieftes Fachwissen zum Schriftspracherwerb, Umgang mit LRS wurde im Rahmen der Zusatzausbildung erworben. Studierende verfügen aber über keine Berufserfahrung im Fach Deutsch auf Primarstufe ^c .
V	Zwischenstufen / fachfremde Grundausbildung <i>n</i> = 12	Studierende hatten eine fachfremde Grundausbildung z.B. in Psychologie, Logopädie, Psychomotorik abgeschlossen und standen mitten im Masterstudiengang in Schulischer Heilpädagogik waren aber noch nicht fertig mit dem Studium. Außerdem verfügten sie entweder über mind. 1-2 Jahre Berufserfahrung im Fach Deutsch auf Primarstufe oder sie standen am Ende der Ausbildung des Masterstudiengangs in Schulischer Heilpädagogik und hatten im Rahmen der Zusatzausbildung als SHP ein vertieftes Fachwissen zum Schriftspracherwerb, Umgang mit LRS erworben.
0	Laien <i>n</i> = 64	Studierende stehen am Anfang des Bachelorstudiengangs zur Grundschullehrperson (2. Semester) oder am Anfang des Masterstudiengangs in Schulischer Heilpädagogik und verfügen über kein Fachwissen zum Schriftspracherwerb und Umgang mit LRS und keine Berufserfahrung im Fach Deutsch auf Primarstufe

^aLRS steht für Lese-Rechtschreibschwäche. ^bSHP steht für Schulische Heilpädagoginnen und Heilpädagogen. ^cStudierende gaben z.B. an in anderen Fächern als Deutsch oder auf Sekundarstufe unterrichtet zu haben.

Aufgrund der geringen Fallzahlen in Substichprobe IV ($n = 9$) und Substichprobe V ($n = 12$) wurden diese beiden Gruppen sowohl aus inhaltlichen Überlegungen wie auch hinsichtlich der ähnlichen Mittelwerte (Anhang, Tabelle 34) mit den anderen Substichproben I, II, III zusammengelegt. Folglich wurden für alle Analysen nur noch drei Substichproben unterschieden. Die Tabelle 9 zeigt, wie sich die Substichproben nach Zusammenlegung hinsichtlich Weiterbildung und Anzahl Jahre Berufserfahrung zusammensetzten.

Tabelle 9: Quantitative Beschreibung der Substichproben in der Hauptstudie

Substichprobe $N = 507$		Weiterbildung in Lesedidaktik		Anzahl Berufsjahre									
				0		1-2		3-5		6-10		>10	
		Zustimmung ^a											
	n	n	%	n	%	n	%	n	%	n	%	n	%
I Primar _{ohne} BE	190	0	0	190	100	0	0	0	0	0	0	0	0
II Primar _{mit} BE	139	27	19	9	7	30	21	38	27	27	19	35	25
III SHP	114	22	19	9	8	15	13	33	29	29	25	28	25
0 Laien	64	0	0	64	100	0	0	0	0	0	0	0	0

^aStudierende gaben an, sich im Bereich Lesedidaktik und/oder zum Umgang mit Lese-Rechtschreibschwierigkeiten weitergebildet zu haben.

Besonderheiten der Laienstichprobe

Sowohl in der Voruntersuchung als auch in der Hauptstudie wurde eine Laienbefragung durchgeführt. Dieser Kontrollgruppenvergleich über eine „Kontraststichprobe“ liefert Hinweise auf die Validität des Instruments und die allfällige Lösbarkeit von Items ohne entsprechendes Professionswissen (Krauss et al., 2017b; Rutsch, 2016). In der Voruntersuchung war die Laienstichprobe aus 22 Einzelpersonen zusammengesetzt, die im privaten Umfeld angefragt worden waren. Bedingung war, dass die Personen einen höheren Bildungsabschluss (Fachhochschule, Höhere Fachschule, Universitäre Hochschule) besaßen und über keinerlei Vorerfahrungen zum Thema Schriftspracherwerb und Umgang mit Leseschwierigkeiten verfügten.

In der Hauptuntersuchung wurde die Laienbefragung bei 40 Studierenden am Anfang des Bachelorstudiengangs zur Grundschullehrperson durchgeführt. Diese Studierenden hatten im Rahmen ihrer Ausbildung noch keinerlei Wissen zum Thema Schriftspracherwerb und Umgang mit Leseschwierigkeiten erworben, und sie hatten noch keine Berufserfahrung. Im Rahmen der Zuordnung zu den Substichproben hatte sich aufgrund der im Kurzfragebogen erfassten Daten gezeigt, dass sich auch unter den Studierenden der übrigen Substichproben weitere 24 Laien befanden, so dass insgesamt 64 Studierende der Laien-Kontraststichprobe zugeordnet wurden.

6 Statistische Methoden

6.1 Testtheoretische Einordnung der Analysen

In der vorliegenden Studie wurden Analysen der Klassischen Testtheorie (KTT) und der probabilistischen Testtheorie (PTT) angewendet, wobei sich die beiden Zugänge gegenseitig ergänzen (Seifert, 2015). Die beiden Testtheorien gehen von unterschiedlichen Annahmen aus, und es werden entsprechend unterschiedliche statistische Methoden zur Verfügung gestellt (Hartig, 2009). Die Kritik an Methoden der KTT sowie konzeptuelle und forschungspraktische Vorteile der PTT machen die Anwendung von Modellen im Rahmen der PTT zunehmend attraktiver (ebd.), was sich in deren Anwendung in zahlreichen neueren Studien zur Kompetenzmessung niederschlägt, u.a. TEDS-M (König, 2012), TEDS-LT (Bremerich-Vos et al., 2011), PISA (Konsortium-PISA, 2009). Im Zusammenhang mit der probabilistischen Testtheorie werden vor allem Modelle der Item-Response-Theorie (IRT) diskutiert, insbesondere das *Rasch-Modell* (Hartig, 2009).

Da in der vorliegenden Untersuchung sowohl Analysen der KTT als auch der IRT durchgeführt wurden, werden die zentralen Eigenschaften der IRT sowie deren Vorteile gegenüber der KTT kurz erläutert und damit die Wahl der angewendeten Methoden begründet.

6.1.1 Eigenschaften der Item-Response-Theorie

Kern der IRT ist der beschriebene „Zusammenhang zwischen beobachtbarem Antwortverhalten und dem dahinterstehenden Persönlichkeitsmerkmal auf Grundlage eines wahrscheinlichkeitstheoretischen Modells“ (Moosbrugger & Kelava, 2012b, S. 420). Modelle der IRT werden deswegen auch als probabilistische Testmodelle bezeichnet (Hartig, 2009).

In IRT-Modellen geht es darum, von beobachtbaren Antworten in einem Test oder Fragebogen auf die Merkmale der Person zu schließen (Hartig, 2009). Im Unterschied zur KTT werden *Antwortmuster* betrachtet, d.h. Antworten auf Items werden als beobachtbare Symptome einer einzigen latenten Variable gesehen (Bühner, 2011). In der IRT wird entsprechend zwischen manifesten und latenten Variablen unterschieden, wobei die manifesten Variablen die beobachtbaren Antworten auf die Items bezeichnen, während es sich bei den latenten Variablen um die Merkmalsausprägungen von nicht beobachtbaren, dahinterliegenden Fähigkeiten handelt, von denen das manifeste Verhalten abhängig ist (Moosbrugger & Kelava, 2012b).

Da in IRT-Modellen die Merkmalsausprägungen und die Schwierigkeit von Items auf derselben Skala beschrieben werden, können umgekehrt auf Basis der individuellen Merkmalsausprägungen Vorhersagen getroffen werden, mit welcher Wahrscheinlichkeit eine Person das Item beantworten kann (Hartig, 2009). Die Wahrscheinlichkeit, ein bestimmtes Item zu lösen, hängt

neben der Fähigkeit der Person von Eigenschaften des Items ab, z.B. der Schwierigkeit, der Ratewahrscheinlichkeit oder der Item-Trennschärfe (Bühner, 2011).

6.1.2 Vorteile der Item-Response-Theorie gegenüber der Klassischen Testtheorie

Die Grenzen der KTT lassen sich gemäß Moosbrugger (2012b) durch Einbezug der IRT überwinden, da die IRT auf einer wesentlich strengeren Annahmebasis beruht als die KTT. Der große Unterschied besteht darin, „dass sie das Reaktionsverhalten der Probanden in Abhängigkeit von Personen- und Itemparametern beschreibt und (...) dem Wahrscheinlichkeitszusammenhang zwischen den Merkmalsausprägungen und dem beobachteten Messwert ausgeht“ (S.116).

Einige Hauptkritikpunkte bezüglich der KTT sollen im Folgenden kurz erläutert und die daraus abgeleiteten Vorteile der IRT aufgezeigt werden:

- Eine Kritik an der KTT liegt darin, dass es sich um eine „Messfehlertheorie“ handelt (Bühner, 2011, S. 53). Es wird angenommen, dass sich ein beobachteter Wert aus einem wahren Wert und einem Messfehler zusammensetzt (ebd.). Diese Annahme ist empirisch nicht überprüfbar, da sowohl der wahre Wert als auch der Fehlerwert nicht direkt beobachtbar sind (Moosbrugger & Kelava, 2012b).
- In IRT-Modellen werden Zusammenhänge zwischen Testverhalten und zu messendem Merkmal mit einem mathematischen Modell beschrieben, was ein konzeptioneller Vorteil gegenüber der KTT ist (Hartig, 2009; Rost, 2004). Das Testverhalten und das zu messende Merkmal werden als separate Größen verstanden. Damit wird die faktische Gleichsetzung von Antworten und Merkmal überwunden (Hartig, 2009).
- Ein Hauptvorteil des zur IRT gehörenden Rasch-Modells ist die Möglichkeit, sowohl Testergebnisse der Probandinnen und Probanden als auch die Schwierigkeiten der Aufgaben auf derselben Skala abzubilden, so dass diese Beziehung für die Interpretation der Testergebnisse genutzt werden kann (ebd.). Kompetenzwerte können damit auch inhaltlich beschrieben werden (Ramseier, 2008).
- Die KTT weist im Bereich der Konstruktvalidität Schwächen auf. Anhand der Modellannahmen besteht keine Möglichkeit zu prüfen, ob die Items bezüglich des untersuchten Merkmals homogen sind, d.h. ob Eindimensionalität wirklich vorliegt (Moosbrugger & Kelava, 2012b). Im Rahmen der KTT muss diese Frage mittels konfirmatorischer Faktorenanalyse geklärt werden (Bühner, 2011).

- Weiter stellt sich das Problem, dass Testkennwerte der KTT wie Itemschwierigkeit, Trennschärfe und Reliabilität stichprobenabhängig sind, und dass sich Werte einer Person in verschiedenen Tests, die dasselbe Konstrukt messen sollten, nicht direkt vergleichen lassen (Bühner, 2011; Moosbrugger, 2012b). Dagegen können bei Gültigkeit eines IRT-Modells Aufgabenschwierigkeiten unabhängig davon bestimmt werden, ob Stichproben mit hohem oder niedrigem Professionswissen untersucht werden (Hartig, 2009).
- Die KTT ist im Prinzip nur für intervallskalierte und kontinuierliche Merkmale uneingeschränkt anwendbar (Bühner, 2011, 54). Die Anwendung auf dichotome Items wie in der vorliegenden Untersuchung erweist sich daher als problematisch (Bühner, 2011). Für IRT-basierte Auswertungen ist die Annahme eines ordinalen Skalenniveaus hingegen nach Hartig (2009) ausreichend.

6.1.3 Das Rasch-Modell

Vorteile des Rasch-Modells gegenüber dem 2PL-Modell (Birnbaum) und dem 3PL-Modell

Die Item-Response-Theorie umfasst verschiedene Modelle, wobei das 1-Parameter-Logistische Modell (1PL-Modell), das sogenannte Rasch-Modell, das einfachste und bekannteste ist (Geiser & Eid, 2010; Hartig, 2009). Es gilt aus verschiedenen Gründen als vorteilhaftestes unter den Latent-Trait-Modellen (Moosbrugger & Kelava, 2012b).

Neben dem 1PL-Modell (Rasch-Modell) sind im Rahmen der IRT das 2PL-Modell (Birnbaummodell) und 3PL-Modell bekannt geworden. Im 2PL-Modell werden im Gegensatz zum Rasch-Modell unterschiedliche Trennschärfen angenommen. Im 3PL-Modell wird zusätzlich zu unterschiedlichen Trennschärfen auch die Ratewahrscheinlichkeit berücksichtigt (Geiser & Eid, 2010). Beide Modelle wären für die Anwendung auf die vorliegenden Daten grundsätzlich interessant, da die Annahme gleicher Trennschärfen in der Praxis oft nicht gegeben ist (Rost, 2004) und bei geschlossenem Antwortformat immer auch richtige Lösungen aufgrund von Raten möglich wären.

Dennoch wurde in der vorliegenden Studie auf die Anwendung mehrparametrischer Modelle verzichtet, da dichotome Items meist nur mit dem Rasch-Modell sinnvoll auf ihre Messeigenschaften geprüft werden können (Bühner, 2011). Für das 2PL- und das 3PL-Modell werden außerdem größere Stichproben benötigt (Bühner, 2011; Hartig, 2009; Rost, 2004). Zudem ist die Ermittlung von Modellparametern im 2PL- und im 3PL-Modell vom statistischen Standpunkt her nicht befriedigend, da sich die zwei bzw. drei Itemparameter nicht unabhängig voneinander schätzen lassen (Rost, 2004, 135).

Grundzüge des Rasch-Modells

Im Folgenden werden die Grundzüge des Modells und die Konsequenzen für die vorliegende Studie dargestellt.

Das Rasch-Modell gehört zu den *Latent-Trait-Modellen*, was sich aus der Idee erklärt, dass beobachtete dichotome oder ordinale Itemantworten messfehlerbehaftete Indikatoren für eine kontinuierliche latente Personenfähigkeit oder Personeneigenschafts-Variable („Latent Trait“) darstellen (Geiser & Eid, 2010, S. 312). Die Latent-Trait-Variable wird meist mit Theta (θ) bezeichnet und bezieht sich im vorliegenden Fall auf das professionelle Wissen/Können von Lehrpersonen im Bereich Lesen (Bühner, 2011; Geiser & Eid, 2010).

Im Rasch-Modell wird angenommen, dass die Wahrscheinlichkeit, dass eine bestimmte Person bei einem Item eine bestimmte Antwort gibt, von nur zwei Parametern abhängt, nämlich 1) von der Schwierigkeit (σ_i) eines Items und 2) vom Wert θ_v der Person auf der latenten Personenfähigkeitsvariable θ (Geiser & Eid, 2010).

Beide Parameter (Itemschwierigkeit und Personenfähigkeit) werden auf einer gemeinsamen Skala („Joint Scale“) abgebildet (Seifert, 2015, S. 134). Die Beziehung zwischen dem Antwortverhalten auf die Items und der Ausprägung auf der latenten Trait-Variable wird in einer mathematischen Gleichung als Itemcharakteristikfunktion („Item Characteristic Curve“ [ICC]) beschrieben (Moosbrugger, 2012a, S.233f.). Diese lässt sich für dichotome Items in folgender Gleichung beschreiben (Geiser & Eid, 2010, S. 313).

$$P(x_{vi} = 1) = \frac{e^{(\theta_v - \sigma_i)}}{1 + e^{(\theta_v - \sigma_i)}},$$

$P(x_{vi} = 1)$ bezeichnet die Wahrscheinlichkeit, dass die Antwort der Person v bei Item i gleich 1 ist
 e bezeichnet die Eulersche Zahl (≈ 2.718)
 θ_v bezeichnet die latente Fähigkeit einer Person
 σ_i bezeichnet die Schwierigkeit eines Items

Itemcharakteristikfunktion im dichotomen Rasch-Modell

Auf der Joint Scale wird für jedes Item eine ICC erstellt, die zeigt, welche Zustimmungswahrscheinlichkeit bzw. Lösungswahrscheinlichkeit bezüglich eines Items in Abhängigkeit von der individuellen Ausprägung einer Person in der latenten Variable zu erwarten ist (Seifert, 2015). Der Zusammenhang wird in der Graphik (Abbildung 17) als logistische Funktion für dichotome Items dargestellt, wobei sich die Enden der Funktion einer Lösungswahrscheinlichkeit von null und eins s-förmig asymptotisch annähern (Bühner, 2011). Die Einheit für den Personen- und Itemparameter ist die Logiteinheit, die einen Wertebereich von minus bis plus unendlich annehmen kann, in der Praxis aber meist zwischen -3 und $+3$ liegt (ebd., S. 496). Ein niedriger Parameterwert zeigt an, dass es sich um ein einfaches Item (Item1) handelt bzw. um eine Person

mit geringer Fähigkeit, während zunehmend positive Wert auf schwierigere Items (Item 3) bzw. auf Personen mit höherer Fähigkeit hinweisen (ebd.). Dabei verschieben sich die Funktionen der Items parallel entlang der σ -Achse, so dass die Items lediglich verschieden schwierig sein können, da nur ein einziger Parameter geschätzt wird (Seifert, 2015).

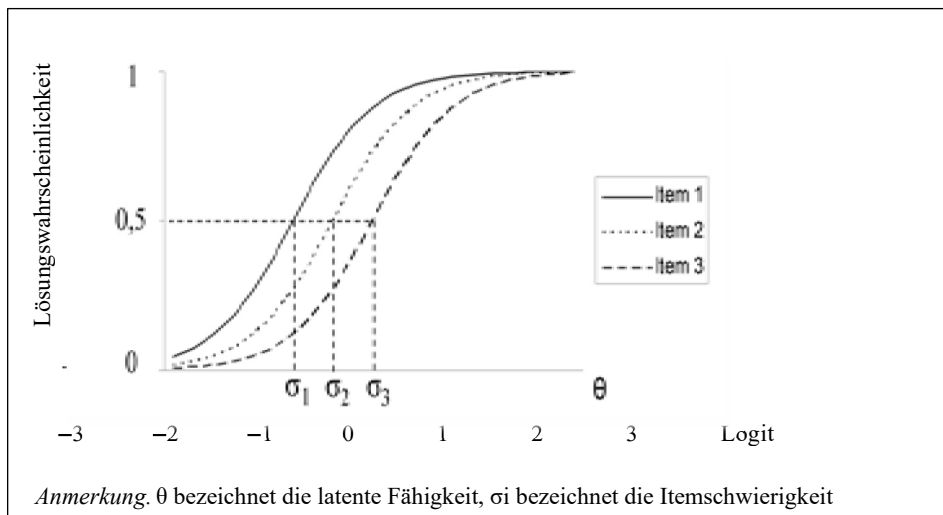


Abbildung 17. Exemplarische Itemcharakteristikfunktion (ICC) im dichotomen Rasch-Modell, ergänzte Graphik in Anlehnung an Geiser & Eid (2010, S. 314) und Bühner (2011, S. 498)

Itemschwierigkeit

In Abbildung 17 lässt sich das Konzept der Itemschwierigkeit im Rasch-Modell erkennen. Die Schwierigkeit ist definiert als der Punkt, an dem die Lösungswahrscheinlichkeit 50% beträgt und der zudem den Wendepunkt der ICC angibt (Geiser & Eid, 2010; Hartig, 2009). Je größer der Wert des Itemschwierigkeitsparameters, umso geringer ist die bedingte Lösungswahrscheinlichkeit für eine bestimmte Person (Eid & Schmidt, 2014). Die Lösungswahrscheinlichkeit eines Items einer Person lässt sich also aus der Differenz des Personenwerts (θ_v) und der Itemschwierigkeit (σ_i) berechnen. Ist die Differenz gleich null, beträgt die Lösungswahrscheinlichkeit 50%, und der Personenwert entspricht genau der Itemschwierigkeit (ebd.).

Der Schwierigkeitsparameter bezeichnet im Rahmen von IRT-Modellen also konzeptionell und interpretativ etwas anderes als im Rahmen der KTT, wo sie dem prozentualen Anteil an richtigen Lösungen eines Items entspricht (Bühner, 2011; Moosbrugger & Kelava, 2012b).

Partial-Credit-Modell

Aufgrund der Kodierung der Daten in der vorliegenden Studie muss an dieser Stelle eine Präzisierung vorgenommen werden. Zu unterscheiden sind das dichotome Rasch-Modell (0-1 Kodierung) und das Partial-Credit-Modell mit teilrichtigen Lösungen bzw. mit unterschiedlich hohen Anforderungen (0-1-2 Kodierung). „Beim Partial-Credit-Modell können dieselben Schätz- und Testmethoden sowie dieselben Itemselektionsstrategien wie beim Rasch-Modell zum

Einsatz kommen“ (Eid & Schmidt, 2014, S. 256). Zusätzlich werden hier Übergänge zwischen den Kategorien (0-1 und 1-2) geschätzt, wobei diese als Schwellen („Thresholds“) bezeichnet werden und mit Hilfe von Schwellenparametern beschrieben sind (Geiser & Eid, 2010, S. 319).

Trennschärfen und Nachweis spezifischer Objektivität

Das Rasch-Modell zeigt weiter, dass die Steigung der ICC immer gleich ist. Das heißt, dass diese Kurven parallel verlaufen (Rost, 2004). Damit wird die dem Rasch-Modell zugrunde gelegte Annahme ausgedrückt, dass kein Trennschärfeparameter angenommen wird bzw. dass alle Trennschärfen auf 1 gesetzt sind (Bühner, 2011). „Dies ist der Grund, weshalb rasch-skalierte Modelle spezifische Objektivität der Vergleiche aufweisen“ (Seifert, 2015, S. 134). Spezifische Objektivität bedeutet, dass der Schwierigkeitsunterschied zwischen zwei Items unabhängig von der Stichprobenszusammensetzung ermittelt werden kann, sofern das Rasch-Modell in der untersuchten Population gilt (Bühner, 2011; Seifert, 2015). Anders ausgedrückt bedeutet dies, dass der Vergleich zweier Personen vom Messinstrument unabhängig ist (Eid & Schmidt, 2014).

Raschhomogenität und lokale stochastische Unabhängigkeit

Eine weitere Annahme des Rasch-Modells ist die lokale stochastische Unabhängigkeit, die besagt, dass Zusammenhänge zwischen Itemantworten einzig durch dieselbe latente Traitvariable erklärt werden, d.h. dass alle Items dasselbe messen (Geiser & Eid, 2010). Bei aller Unterschiedlichkeit in der Formulierung von Items einer Skala sollte also ein inhaltlich homogener Kern berührt werden (Seifert, 2015). Man spricht in diesem Zusammenhang auch von Raschhomogenität einer Skala (Bühner, 2011; Eid & Schmidt, 2014). Lokale stochastische Unabhängigkeit drückt aus, dass „die Antwort eines Probanden auf ein Item [abgesehen von der Fähigkeit] nicht davon abhängt, wie der gleiche Proband andere Items des Tests beantwortet hat“ (Moosbrugger & Kelava, 2012b, S. 278). Zur Gewährleistung stochastischer Unabhängigkeit ist bereits bei der Testkonstruktion auf eine inhaltliche Unabhängigkeit von Items zu achten, etwa indem Items nicht dieselben Fachbegriffe enthalten und mehrere Items nicht auf denselben Itemstamm bezogen sind (Embretson & Reise, 2000).

6.2 Kriterien und Analysen zur qualitativen und quantitativen Beurteilung des Befragungsinstruments

In Anlehnung an Kriterien der FALKO-Studie (Kirchhoff, 2017, S. 129f.) sowie der TEDS-LT-Studie (Stancel-Piatak et al., 2013, S. 36f.) gelten folgende Kriterien für die Itemselektion sowie die Beurteilung des Befragungsinstruments. Die Beurteilungskriterien sind dabei mit den in Kap. 5.1 formulierten Forschungszielen verknüpft.

Qualitative Beurteilung

- Die Items sollen die Facetten des theoretischen Modells angemessen abbilden (Inhaltsvalidität).
- Das Befragungsinstrument soll Themen aufgreifen, die in der Ausbildung und in der Praxis zentral sind.
- Das Befragungsinstrument soll Items enthalten, die nicht aktuelle Trends oder spezifische Lesemethoden (z.B. Lesen durch Schreiben) erfassen.
- Die Itemformulierung und der Kodierleitfaden sollen sowohl theoretisch fundiert sein als auch durch Expertinnen und Experten als inhaltlich adäquat eingestuft werden (Inhaltsvalidität, curriculare Validität).
- Die Kodierung der Items soll objektiv (Interrater-Reliabilität) und ökonomisch sein.

Quantitative Beurteilung

Für einige der im Folgenden genannten Kriterien werden statistisch begründete Beurteilungsrichtlinien bzw. Werte festgelegt (Bühner, 2011, S. 81). Die entsprechenden Werte sind jeweils bei der Beschreibung der angewendeten statistischen Methode in Kap. 6 angegeben. Sie sind grundsätzlich einzuhalten, doch es muss immer auch in Erwägung gezogen werden, in welchen Situationen und wofür ein Instrument später eingesetzt werden soll (ebd.). Insofern müssen für die Itemselektion verschiedene qualitative und quantitative Kriterien aufeinander bezogen und diskutiert werden. Eine a priori Festlegung von festen cut-off-Werten wurde in der vorliegenden Studie aus diesen Gründen nicht als sinnvoll erachtet.

- Die Items sollen ein breites Wissensspektrum angemessen abdecken (Varianz), ohne dass Boden- und Deckeneffekte auftreten (Kap. 6.3.1).
- Die Items sollen hinreichend trennscharf sein und das Konstrukt insgesamt ausreichend reliabel abbilden (Interne Konsistenz). Eine Bewertung der Test-Reliabilität erfolgt aufgrund statistischer Kennwerte (6.3.1) und zudem mit Blick auf inhaltliche Kriterien (Kap. 4).
- Die Gültigkeit des eindimensionalen Rasch-Modells für das Befragungsinstrument soll anhand der In-Fit-Statistiken (Weighted Meansquare, [MNSQ]) überprüft werden (Kap. 6.3.1).

- Das Befragungsinstrument soll hinsichtlich seiner Messinvarianz beurteilt werden. Items mit differenzieller Itemfunktion sollen identifiziert und inhaltlich genauer analysiert werden (Kap. 6.3.3).
- Das Befragungsinstrument soll hinsichtlich seiner theoretisch intendierten Dimensionierung und möglichen alternativen latenten Strukturen beurteilt werden (Kap. 6.3.4).
- Das Befragungsinstrument soll verschiedenen Validitätskriterien genügen: Dazu werden die Ausbildungssensitivität und der Vergleich mit einer Kontraststichprobe geprüft (Kap. 6.3.5).

6.3 Verwendete statistische Verfahren

Im Folgenden werden die statistischen Verfahren erläutert bzw. die oben genannten Beurteilungskriterien weiter spezifiziert. Die statistischen Analysen wurden mit *SPSS* (IBM Corporation, 2017), *ConQuest* (Adams, Wu & Wilson, 2015) und mit *MPlus* (Muthén & Muthén, 2015) durchgeführt. ConQuest eignet sich insbesondere zur Überprüfung psychometrischer Eigenschaften im Rahmen der Anwendung der probabilistischen Testtheorie (Blömeke et al., 2009). Die exploratorische Faktorenanalyse wurde mit *MPlus* gerechnet, was mit *SPSS* nicht möglich war, da es sich im vorliegenden Fall um kategoriale Daten handelt.

6.3.1 Item- und Skalenanalyse in der Klassischen Testtheorie und im Rasch-Modell

Itemschwierigkeit in der KTT

Zur Bestimmung geeigneter Items gehört die Ermittlung der Itemschwierigkeit, d.h. die Identifikation von zu einfachen oder zu schwierigen Items (Bühner, 2011; Moosbrugger & Kelava, 2012b). In der KTT wird bei dichotomen Items (0 = falsche Antwort, 1 = richtige Antwort) als Schwierigkeitsindex (p_i) der relative Anteil an richtigen Lösungen herangezogen (Bühner, 2011). Um Items, die nach dem Partial-Credit-Modell kodiert werden (0 = falsche Antwort, 1 = halbrichtige Antwort, 2 = richtige Antwort), analog interpretieren zu können, wird der Mittelwert durch 2 geteilt (Eid & Schmidt, 2014, S. 131). Wie bei dichotomen Items liegen damit auch diese Itemschwierigkeiten in einem Wertebereich von 0-1.

In der Logik der KTT gemäß Bortz und Döring (2002, S. 218) liegen die bevorzugten Werte in einem mittleren Bereich zwischen $p_i = .20$ - $.80$. Es können aber durchaus auch Mittelwerte in einem Bereich zwischen $p_i = .15$ - $.85$ akzeptiert werden, um eine höhere Schwierigkeitsstreuung zu erreichen und damit ein breites Spektrum an Fähigkeiten in den Extrembereichen zu erfassen (Bühner, 2011). Allerdings können sehr schwierige und sehr einfache Items zu einer reduzierten Homogenität und zu geringen Trennschärfen führen (Bortz & Döring, 2002). In der

vorliegenden Studie waren Items unterschiedlicher Schwierigkeit gewünscht, um das ganze Wissensspektrum auch in den Extrembereichen der bewusst heterogen gewählten Stichprobe abzudecken.

Itemschwierigkeit im Rasch-Modell

Eine anschauliche Darstellung der Itemschwierigkeit im Rasch-Modell bietet die *Map of latent distribution* (Abbildung 18). Sie zeigt die Relationen zwischen Itemschwierigkeit und Personenfähigkeit unter Annahme der Normalverteilung. Dabei werden die Items auf einer Logit-skala mit zunehmender Schwierigkeit von unten nach oben geordnet. „In derselben Rangfolge sind auf der linken Seite der Skala die Personenfähigkeiten aufgetragen, wobei jedes x eine bestimmte Anzahl Fälle repräsentiert“ (Ferber, 2014, S. 52). Unter Geltung des Rasch-Modells können Personen, die einen gleich hohen Logit-Wert aufweisen wie ein bestimmtes Item, dieses Item gemäß Modell mit einer 50%-igen Wahrscheinlichkeit lösen.

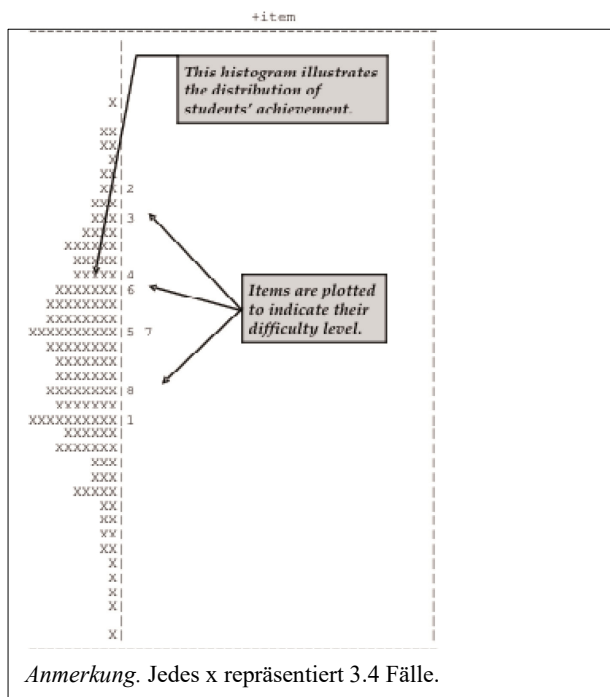


Abbildung 18. Beispiel einer Verteilung der Itemschwierigkeit in Relation zur Personenfähigkeit der latenten Dimension.
ConQuest Manual V2 (Wu, Adams & Wilson, 2007, Version 2.0)

Die Graphik veranschaulicht die Verteilung hinsichtlich der Schwierigkeit für die untersuchte Gruppe und ob ggf. Decken- oder Bodeneffekte bestehen. In *ConQuest* können mehrstufige (z.B. Partial-Credit-Modell) und dichotom kodierte Antworten gemischt werden, d.h. im Falle mehrstufiger Kodierung kann auch die Schwierigkeit der einzelnen Kategorien (0-1-2) für jedes Item angegeben werden.

Schätzmethode von Item- und Personenparameter im Rasch-Modell

Die Schätzung von Itemschwierigkeit und Personenfähigkeit, die aufgrund der Lösungsquote vorgenommen wird, geschieht mittels „Maximum Likelihood Verfahren“ (ML) (Blömeke & König, 2010, S. 255; Rost, 2004, S. 128). In *ConQuest Version 4*, das hier verwendet wird, ist das „Marginal Maximum Likelihood Schätzverfahren“ (MML) implementiert (Heine & Tarnai, 2015, S. 5). Dabei wird ein bestimmter Verteilungstyp angenommen (Rost, 2004, 128). Das Rasch-Modell geht von einer Normalverteilung der latenten Dimension aus. Damit wird die Verteilungsannahme in *ConQuest* festgelegt (Ramseier et al., 2008; Rost, 2004).

Für weiterführende Analysen sind meist die Personenfähigkeiten-Schätzer, die als „Warm Weighted Likelihood Estimates“ (WLE) bezeichnet werden, von Interesse (Bühner, 2011, S. 565). Sie stellen die besten Punktschätzer für Personenparameter dar (ebd.). WLE-Werte haben gegenüber Rohwerten (Summenscores) den Vorteil, dass sie in Abhängigkeit der Itemschwierigkeit und korrigiert um den Messfehler ermittelt werden (Lüdtke & Robitzsch, 2017). Sie werden deshalb in der vorliegenden Arbeit für Gruppenvergleiche verwendet.

Itemtrennschärfe und Itemvarianz in der KTT

Die Trennschärfe gilt in der KTT als wichtigster Kennwert für die Selektion von Items. Inhaltlich drückt die Trennschärfe aus, wie gut ein Item eine Skala, die aus den restlichen Items gebildet wird, widerspiegelt und damit zwischen Personen mit niedriger oder hoher Merkmalsausprägung trennt (Bühner, 2011). Die angestrebten Werte der Item-Skala-Korrelation mit part-whole-Korrektur sollten dabei möglichst hoch ausfallen (ebd.). Für eine genauere Beurteilung der Trennschärfen (r_{it}) wurde in der vorliegenden Studie auf Richtwerte zurückgegriffen, gemäß denen Trennschärfen $< .30$ als niedrig eingestuft werden (ebd., S. 177).

In Bezug auf die Trennschärfen stellt sich hier das Problem, dass sie in einem Verfahren mit hoher Variation der Itemschwierigkeit oft niedrig ausfallen, da schwierige und einfache Items oft eine niedrige Itemvarianz aufweisen, was die Wahrscheinlichkeit einer hohen Trennschärfe wiederum verringern kann (Moosbrugger & Kelava, 2012b).

Items sollten daher nicht einfach aufgrund ihrer Trennschärfe ausgeschlossen werden, sondern primär aufgrund inhaltlicher Überlegungen (Bühner, 2011). Im vorliegenden Fall wurde deshalb ein sehr liberales Selektionskriterium für die Trennschärfe angewendet und eine Untergrenze von $r_{it} = .20$ angelegt.

Die Itemvarianz als weiterer Kennwert gibt die Stärke der Streuung des betreffenden Items an und liefert erste Hinweise auf die Differenzierungsfähigkeit eines Items (Moosbrugger & Kelava, 2012b). Eine hohe Itemvarianz drückt aus, dass sich die Antworten verschiedener

Personen auf ein Item stark unterscheiden (Bühner, 2011). In der Regel hängen Trennschärfe und Itemvarianz dahingehend zusammen, dass eine hohe Varianz positiv mit einer hohen Trennschärfe korreliert (ebd.). Der Zusammenhang zwischen Itemschwierigkeit und Itemvarianz zeigt sich darin, dass mittelschwierige Items in der Regel die höchste Standardabweichung bzw. Varianz aufweisen (ebd.).

Reliabilitätsanalyse – Interne Konsistenz in der KTT

Die Reliabilität – im Sinne interner Konsistenz – gibt an, wie hoch die einzelnen Items durchschnittlich miteinander korrelieren (Döring & Bortz, 2016). Die interne Konsistenz (als eine von vielen Möglichkeiten der Reliabilitätsschätzung) wird dabei meist über den Koeffizienten Alpha (Cronbachs α) berechnet (Bühner, 2011). Konventionell wird die untere Grenze für Cronbachs α meist bei .70 festgelegt (ebd., S. 80). Der Richtwert muss allerdings im Hinblick auf die Heterogenität einer Skala diskutiert werden, und es müssen inhaltliche Entscheide mit einbezogen werden. Denn das Ziel eines Tests kann ja genau darin liegen, Items zu wählen, die inhaltlich vielfältige Aufgaben enthalten, um eine Aussage über ein bestimmtes Konstrukt zu treffen (ebd.). Entsprechend werden auch niedrigere Werte zu erwarten sein (Schermerlell-Engel & Werner, 2012). Ein hoher Koeffizienten sollte auf der anderen Seite nicht als Beleg für Eindimensionalität eines Tests angesehen werden, da die interne Konsistenz auch hoch sein kann, wenn Items mehrdimensionale Merkmale messen (Bühner, 2011; Schermellell-Engel & Werner, 2012).

Weiter führen hohe Trennschärfen neben einer hohen Itemzahl zu einer hohen Reliabilität (Bühner, 2011). Trotzdem ist es nicht immer günstig, aufgrund der Homogenität eine hohe Reliabilität zu erreichen (ebd.). Gerade weil die Schwierigkeit, die Trennschärfe und die Reliabilität zusammenhängen, muss darauf geachtet werden, dass nicht einfach alle Items in den Randbereichen zugunsten der Reliabilität entfernt werden. Durch eine zu starke Homogenisierung kann sich die Inhaltsvalidität verschlechtern (ebd.).

Reliabilität im Rasch-Modell

Auch im Rahmen der Rasch-Skalierung wird die Reliabilität ermittelt, wobei Reliabilität hier ein Maß für die Reproduzierbarkeit der Lageparameter der Items oder Personen bedeutet (Liancre, 2019). „For example, if other persons were given these same items, would the item estimates remain stable?“ (Bond & Fox, 2001, S. 41).

In der Regel werden hier die Item-Separations-Reliabilität („Expected A-Posteriori Estimate/Plausible Value“, [EAP/PV-Reliabilität]) und die Person-Separation-Reliabilität

(„Weighted Likelihood Estimate“, [WLE-Reliabilität]) berichtet (Wu, 2005, S. 117). Die WLE-Reliabilität bezeichnet die geschätzten Personenfähigkeitswerte, die EAP/PV-Reliabilität bezieht sich auf die messfehler- und schätzfehlerbereinigte Testreliabilität der Skala (Stancel-Piatak et al., 2013). Sie gibt die Zuverlässigkeit an, mit der die Verteilung der Fähigkeiten in der Population erfasst wird (Bremerich-Vos & Dämmer, 2013). Die Werte liegen zwischen 0 und 1 und können analog zu Cronbachs α interpretiert werden (Seifert & Schaper, 2012, S. 191). Aufgrund des explorativen Charakters der vorliegenden Studie ist eine Anpassung des Wertes auf .60 allerdings zulässig und schließt damit an den verwendeten Richtwert der TEDS-LT-Studie an (Stancel-Piatak et al., 2013, S. 33).

Mittelwert und Varianz der latenten Personenfähigkeit im Rasch-Modell

Die Gesamtvarianz gilt als weitere Größe zur Einschätzung der Qualität des Befragungsinstrumentes und bezieht sich hier auf die generelle Verteilung um den Mittelwert der latenten Personenfähigkeiten (Neumann, 2014). Sie gibt an, inwiefern durch die Items das Fähigkeitsspektrum der Personen abgedeckt wird und inwieweit das Instrument zwischen fähigen und weniger fähigen Personen unterscheiden kann (ebd.). Wünschenswert ist eine Varianz, die möglichst hoch ausfällt, allerdings gibt es hierzu keine festen Kriterien (Stancel-Piatak et al., 2013). Bei einer Varianz von 1 kann von einem gut variierenden Befragungsinstrument gesprochen werden. Fällt die Varianz gering aus, ist zu prüfen, inwiefern das Instrument Items enthält, die zu einfach oder zu schwierig sind und ob der mittlere Schwierigkeitsbereich nicht hinreichend repräsentiert ist (Neumann, 2014, S. 365; Stancel-Piatak et al., 2013, S. 33). In der TEDS-LT-Studie (Stancel-Piatak et al., 2013, S. 35) wurden Varianzen im Wertebereich um 0.60 als zufriedenstellend klassifiziert. Sie gelten auch in der vorliegenden Studie als Richtwert.

Spezifische Objektivität im Rahmen der Rasch-Skalierung

Die spezifische Objektivität lässt sich mithilfe der gewichteten Abweichungsquadrate („Weighted Mean Square“ [MNSQ]) ermitteln. Damit kann überprüft werden, ob die Annahme paralleler ICCs zutrifft (Seifert, 2015, S. 135). In-Fit-Werte (MNSQ) sollten dabei im Bereich von > 0.80 bzw. < 1.20 liegen, um von einer zufriedenstellenden Passung zum eindimensionalen Rasch-Modell sprechen zu können (König & Blömeke, 2010, S. 255). In anderen Studien werden aber auch In-Fit-Werte von > 0.70 oder < 1.30 akzeptiert (Ramseier, Labudde & Adamina, 2011, S. 16).

Zusätzlich zu den MNSQ-Werten müssen für die Beurteilung der Modellpassung T-Werte berücksichtigt werden. Die T-Statistik prüft, ob die Abweichungen von erwarteten und

beobachteten Häufigkeiten statistisch signifikant sind (Winther, 2010). T-Werte sollten dabei im Bereich -2 bis $+2$ liegen (Konfidenzintervall 95%), um als nicht signifikant abweichend von einer Nullverteilung bei einer Abweichung von nahe 1 ausgehen zu können (Wu & Adams, 2013, S. 352).

6.3.2 Distraktorenanalyse

Bei der Distraktorenanalyse geht es um die Bestimmung geeigneter Distraktoren (Ablenker). Diese sollten weder zu einfach noch zu schwierig sein. Zu einfache Distraktoren werden selbst von Personen mit niedrigem Wissensstand als unplausibel erkannt, und zu „schwierige“ Distraktoren können z.B. widersprüchlich oder von der richtigen Antwort kaum zu unterscheiden sein (Bühner, 2011). Eine Analyse der gewählten Distraktoren *vor* der Umkodierung in das 0-1 Format bzw. das 0-1-2 Format gibt Aufschluss über deren Plausibilität.

Eine gängige Regel zur Analyse von Distraktoren ist die 5%-Regel, wonach Distraktoren von mindestens 5% der Getesteten gewählt werden müssen, um als plausibel zu gelten (Ingenkamp & Lissmann, 2008). In der Pilotstudie wurden Distraktoren, die von weniger als 5% der Probandinnen bzw. Probanden gewählt wurden, um- oder neuformuliert.

In der Hauptuntersuchung wurden die Items nochmals nach demselben 5%-Kriterium analysiert, und als unplausibel identifizierte Distraktoren wurden im Befragungsinstrument markiert. Diese müssen für die weitere Verwendung des Befragungsinstruments überarbeitet werden.

6.3.3 Differenzielle Itemfunktionsanalyse im Rahmen der Item-Response Theorie

Als weitere Methode wird im Rahmen der Testentwicklung zur Validierung von Items die differenzielle Itemfunktionsanalyse durchgeführt (Ferber, 2014; Wilbert & Linnemann, 2011). Ziel der DIF-Analyse ist es, im Rahmen der IRT die Messinvarianz des Instruments zu beurteilen und nachzuweisen, ob in verschiedenen Teilstichproben dieselbe Eigenschaft gemessen wird oder ob Items unterschiedlich funktionieren (Schindler et al., 2019). Zum einen können dadurch Aussagen zur Konstruktvalidität gemacht werden. In der vorliegenden Arbeit soll geprüft werden, ob neben dem angenommenen Professionswissen im Bereich Lesen weitere latente Eigenschaften gemessen werden, wodurch auch die Eindimensionalität in Frage gestellt werden müsste (Wilbert & Linnemann, 2011). Zum anderen können durch die DIF-Analyse Aussagen über die „Fairness“ (ebd., S. 229) gemacht werden, d.h., ob einzelne Items in Subgruppen wesentlich einfacher oder schwieriger zu lösen waren, als aufgrund der Annahme paralleler Schwierigkeitsparameter zu erwarten war. Der Nachweis skalarer Messinvarianz ist insbesondere für den Vergleich von Gruppenmittelwerten bzw. latenten Mittelwerten wünschenswert (Schwab & Helm, 2015). Skalare Messinvarianz liegt vor, wenn neben vergleichbaren

Faktorladungen für jedes Item auf einem Faktor auch die Interzepte für jedes Item über die Gruppen hinweg gleich sind (Gäde, Schermelleh-Engel & Brandt, 2020). Damit sind Schwierigkeitsrelationen der Items und der Nullpunkt der Skala über verschiedene Substichproben vergleichbar. Der Nachweis partieller skalarer Messinvarianz reicht aber für den Vergleich von Mittelwertunterschieden grundsätzlich aus (Christ & Schlüter, 2012).

Eine Möglichkeit DIF zu überprüfen, besteht darin, die postulierte Stichprobenunabhängigkeit zu hinterfragen, d.h. die Stichprobe nach einem relevanten Teilungskriterium in Substichproben zu unterteilen und eine getrennte Itemparameterschätzung vorzunehmen (Moosbrugger, 2012a). In der vorliegenden Studie wurde die Teilung entlang des Ausbildungsniveaus vorgenommen, wodurch ein Vergleich der drei beschriebenen Substichproben (Kap. 5.4.2) möglich wurde.

Ob und bei welchen Items DIF vorliegt, kann sowohl durch einen graphischen Modelltest als auch statistisch überprüft werden. Der graphische Modelltest erfolgt über den paarweisen Vergleich der Itemparameterschätzungen zweier Personen-Subgruppen in einem bivariaten Streudiagramm (Eid & Schmidt, 2014). Weichen die Koordinatenpunkte stark von der Diagonale ab, ist dies ein erster Hinweis für fehlende Messinvarianz bzw. dafür, dass Items in den Subgruppen unterschiedlich funktionieren (Schwab & Helm, 2015). Um drei Substichproben vergleichen zu können, eignet sich eine Graphik, bei der pro Substichprobe für jedes Item die relative Abweichung der Itemschwierigkeit vom Mittelwert 0 abgetragen ist. Um statistisch abzusichern, ob Items in den jeweiligen Gruppen ein problematisches Ausmaß an DIF aufweisen wurden die drei Gruppen jeweils paarweise miteinander verglichen, und jedes Item wurde hinsichtlich DIF überprüft. Für jedes Item wurden Abweichungen vom Ideal der Nulldifferenz berechnet und mittels Gleichung von Draba (1977) statistisch abgesichert (siehe Anhang, Abbildung 42). Gemäß Tristán (2006) muss neben der statistischen Absicherung der DIF ($t > 1.96$, $p < .05$) ein weiteres Kriterium erfüllt sein. Liegt der Betrag der Differenz der jeweils zwei verglichenen Itemschwierigkeiten (Logit) ≥ 0.43 , so kann von einer „slight to moderate“ DIF gesprochen werden. Liegt der Betrag der Differenz der verglichenen Item-schwierigkeiten (Logit) bei ≥ 0.64 ist die DIF als „moderate to large“ einzuschätzen (Tristán, 2006).

Umgang mit differenzieller Itemfunktion

Bei vorliegender DIF – und damit einer Verletzung der Messinvarianz – werden die betreffenden Items normalerweise aus dem Test entfernt (Borsboom, 2006; Tennant & Pallant, 2007). Borsboom (2006) warnt allerdings davor, Items bei vorliegender DIF einfach auszuschließen, da sich dieses Vorgehen negativ auf die inhaltliche Validität eines Verfahrens auswirken kann

und der ausgleichende Effekt verschiedener Item-Bias verschwindet. Das Problem lässt sich also nicht nur statistisch lösen, sondern erfordert einen pragmatischen Ansatz, indem etwa auch der Zweck, für den die Testscores später verwendet werden, einbezogen wird (ebd.). Für Mittelwertvergleiche kann z.B. geprüft werden, inwiefern diese trotz Verzerrungen auf Itemebene robust sind (ebd.). Um die Robustheit von Mittelwertunterschieden auf Basis der WLE-Werte bei vorliegender DIF nachzuweisen, wurden Items zunächst in verschiedene Versionen gesplittet. Damit wurde die Messinvarianz in verschiedenen Modellen operationalisiert, und diese Modelle wurden anschließend verglichen. Diese Alternative zum Ausschluss zahlreicher Items, welche die inhaltliche Validität und Reliabilität des Instrumentes stark einschränken würde, ermöglicht das maximale Ausschöpfen von Items (Du Yates, 1995; Ramseier, 2008; Tennant & Pallant, 2007).

Für das Splitten von Items wurde wie folgt vorgegangen: Wenn bei einem Item der Betrag der Differenz der relativen Itemschwierigkeit zwischen den Gruppen relevant war (moderat oder hoch), wurde es in einem neu erstellten Datensatz so behandelt, als ob es sich bei diesen Gruppen um je eigene, unterschiedliche Items handeln würde, die in der jeweils anderen Gruppe nicht erhoben wurden („missing by design“) (Ramseier, 2008, S. 45). Insgesamt wurden nach diesem Prinzip zwei unterschiedlich gesplittete Datensätze erstellt, deren Mittelwerte (WLE) mit dem Mittelwert ohne Itemsplitting verglichen wurden.

Einen weiteren Umgang mit DIF schlägt Zumbo (2007) vor, wenn er dafür plädiert, DIF nicht nur als Verzerrung oder Test-Unfairness zu sehen, sondern als mögliche Folge unterschiedlicher Antwortprozesse beim Bearbeiten von Items zu betrachten. Die DIF-Analyse ist vor diesem Hintergrund eine Methoden, um zu analysieren, *weshalb* Items von Gruppen unterschiedlich gelöst werden (ebd.). Ein solch interpretativer Umgang mit DIF wird u.a. bei Blömeke et al. (2009) vorgenommen, wo die differenziellen Itemfunktionen in verschiedenen Kohorten vor dem Hintergrund der Art der Lern- und Vergessensprozesse analysiert werden.

6.3.4 Dimensionalität des Befragungsinstruments: Modellierung latenter Variablen und deren Zusammenhänge

Um zu prüfen, ob das Wissen/Können anhand der postulierten *Inhaltsdimensionen* „Diagnostisches Wissen“ und „Grundlagenwissen“ organisiert ist oder ob von einer homogenen Struktur auszugehen ist, wird ein Modellvergleich – eindimensionales versus zweidimensionales Modell – durchgeführt (König & Blömeke, 2010). Die gleiche Fragestellung nach der Dimensionierung stellt sich für die *kognitiven Anforderungen*, die im Modell nach „Fachwissen“ und „fachdidaktischem Wissen/Können“ unterschieden werden (Abbildung 19).

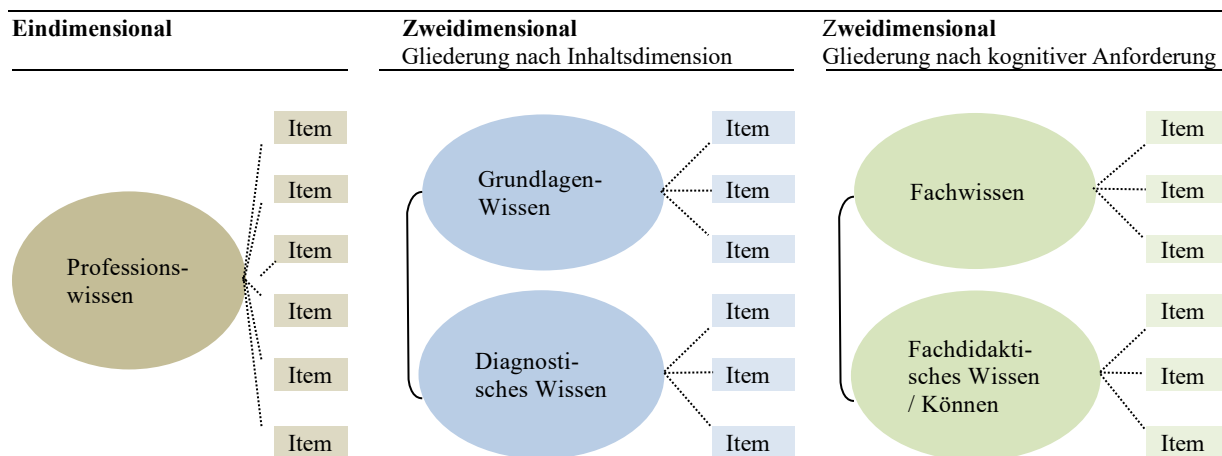


Abbildung 19. Theoretische Modelle, gegliedert nach inhaltlichen Dimensionen und kognitiven Anforderungen

Devianzdifferenz-Test für Modellvergleich „eindimensionales und zweidimensionales Modell“

Um zu prüfen, ob sich die Subdimensionen empirisch unterscheiden lassen, wird analog zum Vorgehen bei HarmoS (Ramseier, 2008; Ramseyer et al., 2008) ein *Devianzdifferenz-Test* durchgeführt. Anhand des Programms *ConQuest* werden mehrdimensionale Modelle geschätzt und mit dem eindimensionalen Modell verglichen (Wu et al., 2007, Version 2.0). Bei diesem Vorgehen zeigt der Devianz-Wert die Abweichung des Modells von den Daten an und kann nur im Vergleich mit der Devianz eines anderen Modells interpretiert werden (Hildebrandt, Jäckle, Wolf & Heindl, 2015; Ramseyer et al., 2008). Um festzustellen, ob ein mehrdimensionales Modell besser passt als ein eindimensionales kann die Devianz anhand der Informationsmasse Bayesian-Information-Criterion (BIC) und Akaike-Information-Criterion (AIC) beurteilt und mittels Devianzdifferenz-Test (χ^2 -Test) statistisch abgesichert werden (Ramseyer et al., 2008; Seifert, 2015). Die Werte weisen darauf hin, „wie wahrscheinlich ein gegebener Datensatz unter Berücksichtigung der Anzahl Parameter des betreffenden Modells ist“ (Rost, 2004, S. 90). Niedrigere BIC- und AIC-Werte zeigen einen besseren Fit des Modells an (ebd.). Da die Devianz eines komplexeren Modells immer geringer ist als jene eines einfacheren Modells würde man allerdings stets das komplexere Modell wählen. Es ist also entscheidend, ob sich die Devianzen bedeutsam voneinander unterscheiden, was mittels Devianzdifferenz-Test (χ^2 -Tests) berechnet werden kann (Grand, Dittrich & Hatzinger, 2017). Ein hoher p -Wert besagt, dass das komplexere Modell (zweidimensionales Modell) gegenüber dem reduzierten Modell (eindimensionales Modell) *keine* bedeutsame Verbesserung bringt. Wird der p -Wert signifikant, ist die Alternativhypothese anzunehmen und dem komplexeren Modell der Vorzug zu geben (Grand et al., 2017).

Exploratorische Faktorenanalyse

Um zu klären, ob anstelle der theoretisch angenommenen Dimensionen alternative latente Dimensionen zugrunde liegen, wird eine exploratorische Faktorenanalyse durchgeführt (Bühner, 2011, 296).

Die Voraussetzungen für die Eignung der Daten für eine Faktorenanalyse werden mittels „Kaiser-Meyer-Olkin-Kriterium“ (KMO-Test) und „Barlett-Test“ geprüft (Bühner, 2011, S. 343f.). Der KMO-Koeffizient gibt an, ob substanzielle Korrelationen zwischen den Items vorliegen und ob diese signifikant von Null abweichen (ebd.). Es gilt die Faustregel, dass der KMO-Wert mindestens .60 betragen muss, um mit der Faktorenanalyse fortzufahren (ebd., S. 347). Der Barlett- χ^2 -Test prüft die globale Nullhypothese, wonach ein signifikantes Ergebnis bedeutet, dass die Korrelationen größer als Null sind und dass sich die Daten grundsätzlich zur Faktorisierung eignen (ebd.).

Um die Anzahl Faktoren zu bestimmen, kann der „Scree-Test“ aufschlussreich sein (Bühner, 2011, S. 343). Hier wird das Extraktionskriterium des Eigenwerts verwendet, das angibt, wie viel der Varianz aller Items durch einen Faktor erklärt wird. Da standardisierte Variablen eine Varianz von 1 aufweisen, bedeutet ein Eigenwert von > 1 , dass der Faktor mehr Varianz erklärt als ein einzelnes Item (Variable) (Moosbrugger & Schermelleh-Engel, 2012). Bei der Extraktionsmethode mittels Scree-Test wird in einem Eigenwertverlauf nach einem bedeutsamen Eigenwertabfall gesucht, was sich in einer entsprechenden Graphik dargestellt lässt. Allerdings handelt es sich dabei nicht um eine objektive Methode (Bühner, 2011).

Eine objektivere Möglichkeit, die Anzahl Faktoren zu bestimmen, liefert der „Likelihood-Quotienten-Test“ und der darin implementierte „ χ^2 -Test“ (ebd., S. 326). Modelle mit unterschiedlicher Faktorenzahl werden gegeneinander getestet, wobei steigende χ^2 -Werte dahingehend interpretiert werden können, dass Modelle mit mehr Faktoren schlechter passen. Da hier die Modellpassung als Nullhypothese formuliert wird, bedeutet ein signifikantes Ergebnis (Signifikanzniveau 5%), dass diese abgelehnt wird und ein Modell mit mehr Faktoren besser passt (ebd., S. 326).

Je nach theoretischer Modellannahme können Faktoren orthogonal (rechtwinklig) oder oblique (schiefwinklig) rotiert werden. Zur Ermittlung möglichst homogener Skalen, die aber auch korreliert sein können, ist in der Regel eine „oblique Rotation“ angezeigt (Bühner, 2011, S. 372). Nach Extraktion der Faktoren müssen die Faktorladungen (λ) beurteilt werden, die das Ausmass der Ausprägung der Itemantwort von einem oder mehreren Faktoren anzeigen (ebd.). Faktorladungen können Werte zwischen +1 und -1 annehmen, wobei als Faustregel gilt, dass Beträge von Ladungen $< .60$ aus dem Test zu entfernen sind (Bortz & Döring, 2002, S. 220).

Die Verwendung von Faktorwerten wird allerdings stark kritisiert, so dass Bühner (2011) dafür plädiert, bei Items mit niedriger Ladung vor der Eliminierung auf die Inhaltsvalidität zu achten. In der vorliegenden Arbeit wurde ein liberaleres Kriterium angewendet. Items mit einer Faktorladung von $\geq .30$ wurden im Itempool behalten, sofern sie eine inhaltlich hohe Bedeutung für das zu messende Konstrukt hatten (Bühner, 2011, S. 371).

Ein weiterer zu beachtender Kennwert zur Bestimmung der Qualität der Items ist die Kommunalität (h^2) (Moosbrugger & Kelava, 2012b, S. 421). Einfach ausgedrückt besagt die Kommunalität, wie gut ein Item durch die Faktoren repräsentiert ist (Bühner, 2011). Sie nimmt Werte im Bereich 0 bis 1 an. Als Richtwert gilt hier, dass Items bei vorliegender Kommunalität von $h^2 < .10$ durch die extrahierten Faktoren nur unzureichend erfasst werden (ebd., S. 358). Gründe für niedrige Kommunalitäten können in der Itemvarianz oder in extremen Itemschwierigkeiten sowie bei schlechten Formulierungen der Items liegen (ebd., 344).

Nicht zuletzt muss zur Interpretation der Dimensionalität einbezogen werden, ob hohe Korrelationen zwischen den Faktoren vorliegen und ob davon ausgegangen werden muss, dass sich die latenten Dimensionen nicht substantiell voneinander unterscheiden und dass eher ein generell übergeordneter Faktor angenommen werden muss (ebd.).

Im vorliegenden Fall wurde für die exploratorische Faktorenanalyse das Programm *MPlus* (Muthén & Muthén, 1998-2012) verwendet, das im Gegensatz zu *SPSS* die Möglichkeit bietet auch mit dichotomen oder ordinalen Daten eine Faktorenanalyse durchzuführen (Bühner, 2011). Dabei wurden die Variablen als „kategorial“ spezifiziert, da es sich um ordinale Werte handelt. Als Schätzmethode wurde das ML-Verfahren verwendet und die Einstellung „oblique Rotation“ gewählt (Geiser, 2010).

6.3.5 Validierung: Ausbildungssensitivität und Kontrastgruppenvergleich

Im Anschluss an die Skalierung des Instruments und an Analysen zur Datenstruktur wurden zur Validierung des Verfahrens Vergleiche zwischen den gebildeten Substichproben durchgeführt. In Anlehnung an das Vorgehen bei Rutsch (2016) und Seifert (2015) wurde der querschnittliche Wissensanstieg im Bereich Lesen über verschiedene Ausbildungsstufen (Substichproben) hinweg betrachtet und zudem ein Vergleich mit der Kontraststichprobe der Laien durchgeführt.

Datengrundlage für die Mittelwertvergleiche

Als Datengrundlage dienten die im Rahmen der Rasch-Skalierung ermittelten WLE-Werte der Skala mit 18 Items, die für das finale Befragungsinstrument einbezogen wurden.

Auch für die Laienstichprobe wurden im Anschluss an die Skalierung des Befragungsinstruments an den drei Substichproben WLE-Werte ermittelt und für den Kontrastvergleich

verwendet. Um die Fähigkeiten der Laien im Vergleich zu denen der Hauptstichprobe erfassen zu können, wurde die in der Hauptstichprobe festgelegte Skala als bekannt eingeführt, und die WLE-Werte für die Laien wurden neu geschätzt.

Aufgrund der fehlenden Messinvarianz (DIF) musste für weitere Analysen entschieden werden, welche WLE-Werte zugrunde gelegt werden sollten, da sich diese unter Verwendung verschiedener Itemsplitting-Prozeduren unterscheiden. Für den Vergleich der Substichproben und den Kontrastgruppenvergleich wurden die WLE-Werte des Basismodells ohne Itemsplitting eingesetzt. Eine genauere Begründung dafür findet sich in den Ergebnissen in Kap. 7.2.3.5.

Voraussetzungen: Verteilung der Skalenwerte

Für den Vergleich des Wissens/Könnens zwischen den verschiedenen Substichproben und der Laien-Kontraststichprobe wurde die Verteilung der Werte hinsichtlich Normalverteilung und Varianzhomogenität getestet (Eid, Gollwitzer & Schmitt, 2015). Hier musste zum einen die Verteilung der erreichten Punkte (Summenscores) wie auch der Personenfähigkeits-Schätzer (WLE) hinsichtlich ihrer Abweichung von einer Normalverteilung mittels „Kolmogoroff-Smirnoff-Test“ untersucht werden (Bühner & Ziegler, 2009, S. 94f.). Die Varianzhomogenität wurde mittels „Levene-Test“ überprüft (Eid et al., 2015, S. 409).

Neben der Verteilung der Werte wurde auch die Übereinstimmung der WLE-Werte mit den Summenscores berechnet (Korrelationskoeffizient nach Pearson) (Bortz & Schuster, 2010). Zu erwarten war ein hoher Zusammenhang der beiden Variablen.

Mittelwertvergleiche der Substichproben und Effektstärken

Zur Überprüfung der Kriteriumsvalidität wurden nach dem Kriterium des Ausbildungshintergrunds drei Substichproben gebildet, die hinsichtlich ihres Wissenszuwachses verglichen wurden (Pissarek & Schilcher, 2017). Es wurde angenommen, dass sich das Professionswissen der Personen der Substichproben primär aufgrund ihrer Ausbildungsbiografie und ihrer Berufserfahrung unterscheidet. Erwartet wurde, dass Studierende mit einer Zusatzausbildung in Schulischer Heilpädagogik höhere Wissenswerte erreichen als Studierende ohne Berufserfahrung, die am Ende der Grundausbildung stehen, oder Studierende, welche die Grundausbildung abgeschlossen haben und zusätzlich über Berufserfahrung verfügen.

Zur Analyse von Mittelwert-Gruppenunterschieden wurde zunächst eine einfaktorielle Varianzanalyse (ANOVA) (Bühner & Ziegler, 2009). Um die Unterschiede der einzelnen Gruppen zu bestimmen, wurde anschließend ein post-hoc Einzelvergleich durchgeführt. Im Unterschied zu a-priori-Kontrasten, bei denen bereits vor der Testung Hypothesen über verschiedene

Gruppenmittelwerte bestehen, dienen post-hoc-Tests der Hypothesengenerierung (ebd.). Hier werden in der Regel der „Bonferroni-Test“ oder „Scheffé-Test“ verwendet (ebd., S. 546). Beide Verfahren eignen sich für Gruppenvergleiche mit homogenen Varianzen (ebd.). Bei ungleichen Varianzen wird alternativ der „Games-Howell-Test“ vorgeschlagen (Field, 2013, S. 459).

Zur Absicherung der Mittelwertunterschiede wurde das Effektstärkemaß η^2 (η^2) verwendet (Bühner & Ziegler, 2009). Es gibt an, wie groß der Anteil der abhängigen Variable an der Gesamtvariabilität ist, der durch die Gruppenunterschiede erklärt wird (ebd.). Zur Interpretation liegen die Konventionen von Cohen (1988) vor, wonach $\eta^2 > 0.01$ einen kleinen Effekt, $\eta^2 > 0.06$ einen mittleren Effekt und $\eta^2 > 0.14$ einen starken Effekt bezeichnet (Döring & Bortz, 2016, S. 821).

Für die post-hoc Einzelvergleiche wurde das Effektstärkemaß d nach Cohen (1988) verwendet, das sich aus der Mittelwertdifferenz der zu vergleichenden Gruppen, dividiert durch die gepoolte Standardabweichung, ergibt (Pissarek & Schilcher, 2017). In der Klassifikation von Cohen (1988) entspricht eine Effektstärke von $d = 0.20$ einem kleinen, von $d = 0.50$ einem mittleren und von $d = 0.80$ einem starken Effekt (Döring & Bortz, 2016, S. 820).

Die Voraussetzungen normalverteilter Werte pro Substichprobe sowie die Varianzhomogenität wurden mit oben erwähnten Verfahren im Vorfeld geprüft, auch wenn die Varianzanalyse bei Stichproben $N > 30$ gegenüber Verletzungen der Normalverteilung robust ist (Bühner & Ziegler, 2009; Eid et al., 2015). Bei Verletzung der Varianzhomogenität können robuste Verfahren wie der „Brown-Forsythe-Test“ oder „Welch-Test“ angewendet werden (Eid et al., 2015, S. 409).

Mittelwertvergleich mit Kontraststichprobe

Zur Überprüfung der Validität wurde ein Kontrastgruppenvergleich durchgeführt. Dieses Verfahren wurde z.B. bei Rutsch (2016 131) und bei COACTIV (Krauss et al., 2011) im Rahmen der Validierung der Kompetenz- bzw. Wissenstests angewendet. Damit wird untersucht, inwiefern das Verfahren tatsächlich *fachspezifisches* Wissen von Lehrpersonen abfragt und nicht beiläufig erworbenes Wissen.

In der vorliegenden Studie wurde angenommen, dass die Laien signifikant schlechter abschneiden als die Probandinnen und Probanden der übrigen Substichproben. Diese Annahme wurde ebenfalls mit dem oben genannten post-hoc-Vergleich überprüft, wiederum basierend auf den WLE-Werten. Für die Analyse wurden die beiden oben beschriebenen Verfahren – einfaktorielle Varianzanalyse und post-hoc Einzelvergleich – durchgeführt. Auch die Effektstärke wurde nach demselben Verfahren wie für Mittelwertvergleiche der Substichproben berechnet.

7 Ergebnisse

7.1 Ergebnisse der Voruntersuchung

In der Voruntersuchung wurde ein Entwurf mit 24 Items getestet. Die betreffenden Items wurden aus einem Itempool von anfänglich 56 Items ausgewählt und an einer Stichprobe von $n = 179$ bzw. $N = 201$ (inklusive Laien) pilotiert. Im Einzelnen setzte sich die Stichprobe – wie später auch die Hauptuntersuchung – formal aus vergleichbaren Substichproben zusammen. Um eine begründete Auswahl von Items zu treffen, die später für die Hauptstudie eingesetzt werden konnten, wurde das Befragungsinstrument bereits in der Voruntersuchung mit einer kleinen Stichprobe hinsichtlich statistischer Merkmale beurteilt. Dabei wurden aber nicht alle Analysen der Hauptuntersuchung durchgeführt, sondern es wurde lediglich eine erste Einschätzung vorgenommen. Der folgende Überblick (Tabelle 10) zeigt, welche Analysen an welcher Stichprobe durchgeführt wurden.

Tabelle 10: Substichproben und Übersicht der durchgeführten Analysen der Voruntersuchung

Substichproben	n	Item- und Skalenanalyse	Dimensionalität	Validierung Ausbildungssensitivität und Kon- trastgruppenvergleich
I Primar _{ohne} BE	66	x	x	x
II Primar _{mit} BE	61	x	x	x
III SHP	43	x	x	x
IV SHP Zwischenstufe ^a	9	x	x	–
0 Laien	22	–	–	x
Σ	201	179	179	192

^aDie SHP hatten das Studium noch nicht abgeschlossen und hatten noch nicht alle Inhalte vermittelt bekommen.

Für die Validierung des Befragungsinstruments wurden die neun zur „Zwischenstufe“ gehörigen Fälle nicht einbezogen, um möglichst unterscheidbare Substichproben für den Mittelwertvergleich zu erhalten.

7.1.1 Item- und Skalenanalyse

Die folgenden Analysen basieren einerseits auf der KTT, andererseits werden Ergebnisse auf Grundlage des Rasch-Modells referiert.

Analysen auf Grundlage der KTT

In einem ersten Schritt wurden 24 verwendbare Items aufgrund niedriger Trennschärfen ($r_{it} < .20$) um acht Items reduziert. Ein weiteres Item wurde aufgrund seiner extrem hohen Lösbarkeitsrate ($p_i = .87$) aus dem Itempool entfernt. Für die verbleibenden 15 Items ergeben sich damit auf Grundlage der KTT folgende statistische Kennwerte (Tabelle 11).

Tabelle 11: Statistische Kennwerte als Grundlage für die Itemanalyse in der Voruntersuchung mit 15 Items

Item Nr. ^a	Kodierung	Schwierigkeitsindizes p_i	Itemvarianz σ	part-whole korrigierte Trennschärfe r_{it}
3	Partial-Credit	.57	0.53	.29
4	Partial-Credit	.79	0.38	.26
5	0-1 Kodierung	.86	0.13	.28
6	0-1 Kodierung	.78	0.18	.24
7	0-1 Kodierung	.23	0.18	.30
8	0-1 Kodierung	.35	0.23	.37
9	0-1 Kodierung	.60	0.24	.27
10	0-1 Kodierung	.35	0.23	.29
11	0-1 Kodierung	.15	0.13	.42
14	0-1 Kodierung	.81	0.16	.25
19	Partial-Credit	.50	0.52	.39
20	Partial-Credit	.62	0.41	.29
21	Partial-Credit	.81	0.44	.32
22	0-1 Kodierung	.27	0.20	.46
23	0-1 Kodierung	.18	0.15	.37
Max. 20 Punkte				Cronbachs $\alpha = .70$

^aNr. der Items aus der Voruntersuchung entsprechen nicht denjenigen in der Hauptuntersuchung.

Die verbleibenden 15 Items zeigen befriedigende Itemschwierigkeiten. Cronbachs α liegt knapp über dem akzeptablen Wert. Die Itemvarianzen sowie die Itemtrennschärfen weisen allerdings in einigen Fällen eher niedrige Werte auf. Bezüglich Itemvarianzen zeigt sich ein Vorteil von Partial-Credit-Kodierungen gegenüber den zweistufigen Kodierungen.

Analysen auf Grundlage des Rasch-Modells

In einem nächsten Schritt wurden die Items auf Grundlage des Rasch-Modells skaliert, zunächst unter Einbezug aller 24 Items und anschließend unter Ausschluss der neun Items, welche die Kriterien nach KTT nicht erfüllten. In beiden Fällen wiesen die Items einen In-Fit (MNSQ) zwischen > 0.80 und < 1.20 auf und passten damit zum eindimensionalen Rasch-Modell (Wu & Adams, 2013). Die Varianz für 15 Items liegt bei 0.78, die EAP-Reliabilität bei 0.73 und die WLE-Reliabilität bei 0.71. Diese Werte können insgesamt als befriedigend eingestuft werden. Auch die Itemschwierigkeiten, die in der *Map of latent Distribution* (Abbildung 20) als Relation zwischen Item- und Personenparametern visualisiert sind, liegen weitestgehend in einem zufriedenstellenden Bereich.

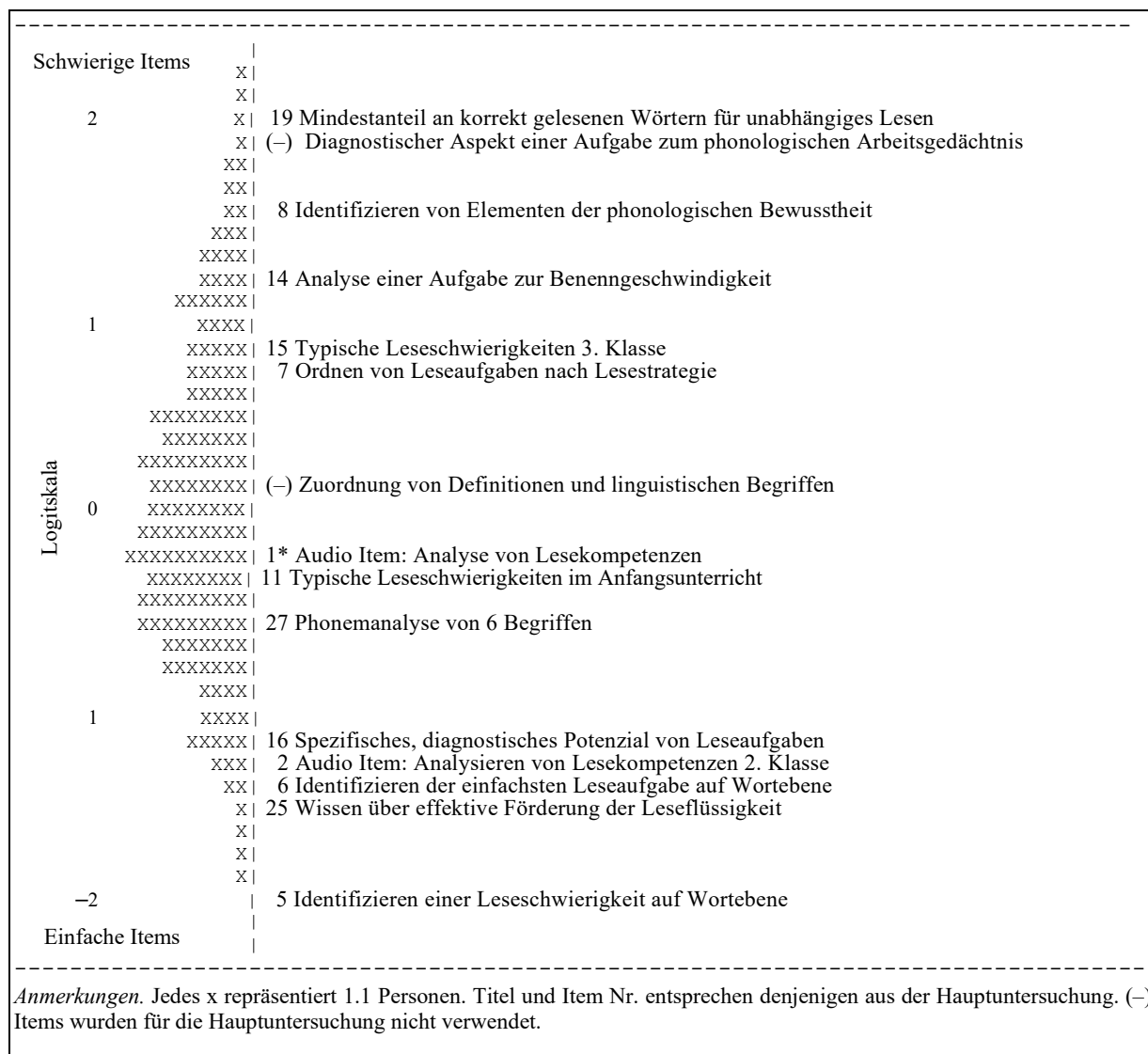


Abbildung 20. Map of latent Distribution: Verteilung der Itemschwierigkeit (rechts) in Relation zur Verteilung der Personenfähigkeit (links) in der Voruntersuchung

Insgesamt zeigt sich eine breite und gleichmäßige Verteilung der Schwierigkeitsgrade. Einzig Item 5 fällt als besonders einfaches Item auf. Dieses Item war allerdings „als warm-up“-Frage konzipiert und wurde aus inhaltlichen und motivationalen Gründen trotz seiner hohen Lösungswahrscheinlichkeit im Itempool für die Hauptuntersuchung belassen.

Zwei Items wurden trotz zufriedenstellender statistischer Kennwerte in der Voruntersuchung für die Hauptstudie ausgeschlossen, und zwar das Item „*Diagnostischer Aspekt einer Aufgabe zum phonologischen Arbeitsgedächtnis*“ sowie das Item „*Zuordnung von Definitionen und linguistischen Begriffen*“. Beide Items wurden im Rahmen der zweiten Validierung mit Expertinnen und Experten durch neue Items ersetzt.

Item 1, 15 und 25 erreichten in der Hauptstudie den kritischen Trennschärfewert nicht mehr und wurden für das Befragungsinstrument der Hauptstichprobe ausgeschlossen. Dies wird damit begründet, dass bestehende Items aus der Voruntersuchung teilweise zu Gunsten

eindeutigerer Antwortalternativen überarbeitet wurden, was in manchen Fällen aber zu schlechteren Item-Skalen-Kennwerten geführt hat. Deutlich wird dies bei Audio-Item 3 (entspricht Audio-Item 1 der Hauptstudie) (Abbildung 21), das durch die Reduzierung der Distraktoren sowie Änderungen im Wortlaut von einer Trennschärfe von $r_{it} = .29$ auf eine unbefriedigende Trennschärfe von $r_{it} = .08$ abgefallen ist.

Item in der Voruntersuchung

Das Kind steht am Ende der 2. Klasse . Kreuzen sie alle Kompetenzen an, die das Kind mehrheitlich zeigt. Machen sie zum Schluss eine Gesamteinschätzung (g).		
Alter: 8 Jahre 9 Monate Stufe: Ende 2. Klasse (Juni)		
a	Synthetisierendes gedehntes Lesen	<input type="checkbox"/>
b	Nutzen von größeren Einheiten (z.B. Silben)	<input type="checkbox"/>
c	Sichtwortschatz bei kurzen Wörtern (bis 5 Buchstaben)	<input type="checkbox"/>
d	Sichtwortschatz bei längeren/komplexeren Wörtern (ca. mehr als 5 Buchstaben)	<input type="checkbox"/>
e	Lesegenauigkeit	<input type="checkbox"/>
f	Sinnvolle Betonung / Prosodie	<input type="checkbox"/>
g	Gesamteinschätzung: Die Lesekompetenz ist altersentsprechend, unauffällig.	Ja <input type="checkbox"/> Nein <input type="checkbox"/>

Überarbeitetes Item in der Hauptstudie

Das Kind steht am Ende der 2. Klasse . Kreuzen Sie alle Kompetenzen an, die das Kind mehrheitlich zeigt. (Mehrere Antworten ankreuzen) Nehmen Sie zum Schluss eine Gesamteinschätzung vor (e). (Eine Antwort ankreuzen)		
Alter: 8 Jahre 9 Monate Stufe: Ende 2. Klasse (Juni)		
a	Lesefehler und Auslassungen werden mehrheitlich selbst korrigiert.	<input type="checkbox"/>
b	Größere Einheiten (z.B. Silben) werden mehrheitlich genutzt.	<input type="checkbox"/>
c	Sichtwortschatz bei kurzen Wörtern (bis 5 Buchstaben) mehrheitlich vorhanden.	<input type="checkbox"/>
d	Sichtwortschatz bei längeren/komplexeren Wörtern (mehr als 5 Buchstaben) mehrheitlich vorhanden.	<input type="checkbox"/>
e	Gesamteinschätzung: Die Ziele für das Lesen am Ende der 2. Klasse sind...	übertroffen <input type="checkbox"/> erreicht <input type="checkbox"/> noch nicht erreicht <input type="checkbox"/>

Abbildung 21. Beispiel Audio-Item 1 vor und nach der Überarbeitung in der Voruntersuchung und Hauptstudie

7.1.2 Dimensionalität des Befragungsinstruments

Ein erster Hinweis zur Dimensionalität des Befragungsinstruments wurde mittels exploratorischer Faktorenanalyse gewonnen. In der Voruntersuchung wurde die theoretische Struktur des Modells lediglich indirekt überprüft, indem die Faktorenstruktur und die theoretisch zugeordneten Items verglichen wurden. Eine hypothesenprüfende Analyse der postulierten Strukturen mittels Devianz-Test erfolgte erst mit den Daten der Hauptuntersuchung.

Die Items erfüllen die Voraussetzungen für eine Faktorenanalyse gemäß Kaiser-Meyer-Olkin - Kriterium ($KMO = .75$) und Barlett-Test ($\chi^2(120) = 370.93, p < .001$). In der folgenden Tabelle

12 wird das Modell mit zwei Faktoren unter Verwendung der obliquen Rotation und ML-Schätzer für ordinale Daten unter Einbezug von 15 Items für die Stichprobe ohne Laien aufgezeigt. Dabei werden die Items zunächst entlang den Dimensionen *Grundlagenwissen* und *Diagnostisches Wissen* und im zweiten Schritt den Dimensionen *Fachwissen* und *fachdidaktisches Wissen/Können* zugewiesen.

Tabelle 12: Dimensionalität des Befragungsinstruments in der Voruntersuchung

Item Nr. ^a	Dimension		Faktorladung λ		Kommunalität
	Diagnostisches Wissen (DW)/Grundlagenwissen (GW)	Fachwissen (FW)/Fachdi- daktisches Können (FDK)	Faktor 1	Faktor 2	h^2
3	DW	FDK	.29	.15	.15
4	DW	FDK	-.19	.75*	.48
5	DW	FDK	.29	.36*	.30
6	GW	FDK	.57*	-.09	.29
7	GW	FW	.61*	-.12	.32
8	GW	FDK	.64*	.01	.41
9	DW	FW	.30*	.19	.18
10	DW	FW	.01	.63*	.40
11	GW	FW	.65*	.29	.66
14	GW	FDK	.48*	.01	.23
19	GW	FW	.30*	.24	.21
20	GW	FDK	.19	.23	.13
21	DW	FDK	.05	.61*	.40
22	DW	IFW	.81*	.13	.76
23	DW	FW	.75*	-.06	.53
Eigenwert			4.78	1.67	
% der kumulierten Varianz			31.88	42.94	

Anmerkung. Die beiden Faktoren korrelieren signifikant, $r(179) = .44, p < .05$. ^aNr. der Items in der Voruntersuchung sind nicht identisch mit denjenigen der Hauptuntersuchung. * $p < .05$.

Es fällt zum einen auf, dass auf dem ersten Faktor deutlich mehr Items laden als auf dem zweiten. Der Anteil erklärter Varianz des ersten Faktors liegt deutlich über dem des zweiten Faktors. Die Kommunalitäten von Item 3 und Item 20 sind zudem sehr niedrig, so dass diese Items bei einer Zweifaktorenlösung ausgeschlossen werden müssten. Liegt die Kommunalität bei $h^2 < .10$, werden diese Items nur unzureichend von den extrahierten Faktoren repräsentiert. Weiter wird deutlich, dass sich die Unterscheidung zwischen den Dimensionen *Fachwissen* und *fachdidaktisches Können* empirisch nicht zeigt. Auch die zweifaktorielle Struktur entlang den Dimensionen *Diagnostisches Wissen* und *Grundlagenwissen* konnte nicht gefunden werden. Da die beiden Faktoren substanziell miteinander korrelieren und theoretisch auch einige Gründe für eine eindimensionale Skala sprechen, wurde im Weiteren eine einfaktorielle Lösung geprüft. Hier zeigt sich, dass alle 15 Items bei der einfaktoriellen Lösung signifikant auf einen Faktor laden (5% Signifikanzniveau) und in einem Bereich von $\lambda = .35 - .85$ liegen. Auch hier liegen die Kommunalitäten für Item 3, 4, 9, 15, 20 teilweise nur knapp über dem akzeptablen Bereich von $h^2 > .10$. Die Inspektion des Scree-Tests mittels Eigenwertkriterium (Anhang, Abbildung 41) lässt zudem die Interpretation einer Einfaktorenlösung zu.

Aufgrund dieser Analysen scheint es plausibel 15 Items, die eine einfaktorielle Struktur abbilden und zudem Trennschärfen von $r_{it} > .20$ aufweisen, für die Hauptstudie zu übernehmen. Die Items liegen sowohl nach klassischer als auch nach probabilistischer Testtheorie in einem angemessenen Schwierigkeitsbereich und decken ein breites Wissensspektrum ab.

7.1.3 Validierung des Befragungsinstruments

Datengrundlage für die Validierung des Befragungsinstruments in der Voruntersuchung waren die Summenscores (erreichte Punkte) der reduzierten Stichprobe ($N = 192$). Zur Validierung wurde zum einen die Ausbildungssensitivität geprüft, zum anderen wurde der erwartete Wissensunterschied zur Kontraststichprobe der Laien untersucht.

Für die ANOVA und post-hoc-Einzelvergleiche wurden entsprechende Voraussetzungen getestet. Dabei zeigte der Kolmogorov-Smirnov-Test, dass die Summenscores in den einzelnen Substichproben mit Ausnahme der Kontraststichprobe der Laien signifikant (Signifikanzniveau 5%) von einer Normalverteilung abweichen. Allerdings wirkt sich eine Verletzung der Normalverteilung nicht wesentlich auf die Ergebnisse der ANOVA aus. Der Levene-Test zeigte das Vorliegen von Varianzhomogenität der vier Gruppen ($p = .95$). Der Haupteffekt der einfaktoriellen ANOVA für gleiche Varianzen wurde unter Einbezug der Laien signifikant, $F(3, 188) = 59.86$, ($p < .01$, $N = 192$) bei einer Effektstärke von $\eta^2 = 0.489$.

Der querschnittliche Wissensanstieg der Summenscores (max. Punktzahl = 20) zeigte sich in erwarteter Richtung: von Laien ($M = 5.95$, $SD = 2.57$, $n = 22$) über $\text{Primar}_{\text{ohne BE}}$ ($M = 9.15$, $SD = 2.64$, $n = 66$) zu $\text{Primar}_{\text{mit BE}}$ ($M = 10.98$, $SD = 2.78$, $n = 61$) und zur kompetentesten Gruppe der SHP ($M = 14.65$, $SD = 2.82$, $n = 43$). Damit liegt ein wichtiger Hinweis vor, dass die 15 einbezogenen Items Wissensunterschiede zwischen den Substichproben abbilden können. Wie sich dies im Einzelnen darstellt, zeigt der post-hoc-Einzelvergleich mit *Bonferroni-Korrektur* (Tabelle 13).

Tabelle 13: Gruppenunterschiede zwischen Substichproben und Laien-Kontraststichprobe, Bonferroni Test

Einzelvergleich der Substichproben		Mittlere Differenz des Summenscore	p	95%-Konfidenzintervall		Effektstärke d
				unten	oben	
0 Laien	I $\text{Primar}_{\text{ohne BE}}$	3.20	.000	1.41	4.98	1.22
I $\text{Primar}_{\text{ohne BE}}$	II $\text{Primar}_{\text{mit BE}}$	1.83	.001	0.54	3.12	0.67
II $\text{Primar}_{\text{mit BE}}$	III SHP	3.67	.000	2.22	5.11	1.31

Alle Mittelwertunterschiede zwischen den Gruppen sind signifikant (Signifikanzniveau 5%). Die Effektstärken (Cohens d) liegen in einem mittleren ($d = 0.50$) bis hohen ($d = 0.80$) Bereich (Döring & Bortz, 2016, S. 820).

Somit hatte die Voruntersuchung bestätigt, dass sich das beabsichtigte Vorgehen zur Entwicklung eines ausbildungssensitiven Instruments grundsätzlich eignete, und dass für die Hauptuntersuchung in diese Richtung weitergedacht werden konnte.

7.2 Ergebnisse der Hauptstudie

Aufgrund der Ergebnisse der Voruntersuchung wurden 15 Items beibehalten. Wegen niedriger Trennschärfen wurden sie jedoch teilweise stark überarbeitet. Damit das theoretische Konstrukt des Modells durch die Items weiterhin angemessen abgebildet werden konnte, wurden zahlreiche Items nachkonstruiert und zunächst durch qualitative Validierungsprozesse (Kap. 5.3.3) schrittweise reduziert. In der Hauptstudie lagen schließlich 28 Items vor, die den beschriebenen statistischen Analysen unterzogen wurden. Die Ergebnisse der Analysen werden in der Reihenfolge ihrer Erwähnung (Kap. 6) detailliert dargestellt.

Der folgende Überblick (Tabelle 14) zeigt, welche Analysen an welcher Stichprobe durchgeführt wurden. Zudem ist festgehalten, wie viele Items jeweils in die Analyse einbezogen wurden. Es schien teilweise sinnvoll, das Ergebnis unter Einbezug aller 28 Items (Skala₂₈) mit dem Ergebnis des auf 18 Items (Skala₁₈) reduzierten Befragungsinstruments zu vergleichen. In gewisse Analysen wurden aber nur noch die 18 Items mit befriedigenden statistischen Kennwerten aus der Item- und Skalenanalyse mit einbezogen. Die Distraktorenanalyse schloss ausserdem nur Items mit gebundenen Formaten ein.

Tabelle 14: Substichproben und Übersicht der durchgeführten Analysen der Hauptstudie

Substichproben	n	Item- und Skalenanalyse	Distraktoren- analyse	DIF-Analyse	Dimensionalität	Validierung: Ausbildungssensitivität Kontrollgruppenvergleich
		28/18 Items	21 Items	18 Items	28/18 Items	18 Items
I Primar _{ohne} BE	190	x	x	x	x	x
II Primar _{mit} BE	139	x	x	x	x	x
III SHP	114	x	x	x	x	x
0 Laien	64	–	x	–	–	x
Σ	507	443	507	443	443	507

7.2.1 Item- und Skalenanalyse

7.2.1.1 Itemtrennschärfe, Itemvarianz und Reliabilität in der Klassischen Testtheorie

Im Rahmen der Item- und Skalenanalyse wurden verschiedene Analysen durchgeführt, die in Tabelle 15 dargestellt sind. Die Analysen betreffen die Skala₂₈ sowie Skala₁₈ und identifiziert Items, die ungünstige statistische Kennwerte erzielt haben.

Tabelle 15: Statistische Kennwerte als Grundlage für die Itemanalyse in der Hauptuntersuchung

Item Nr.	Kodierung	Schwierigkeitsindizes p_i	Itemvarianz σ	Skala ₂₈	Skala ₁₈
				part-whole korrigierte Trennschärfe r_{it}	
1	Partial-Credit	.68	0.51	.08	–
2	Partial-Credit	.75	0.49	.25	.23
3	Partial-Credit	.61	0.62	.18	–
4	Partial-Credit	.61	0.31	.14	–
5	0-1 Kodierung	.81	0.15	.33	.30
6	0-1 Kodierung	.72	0.20	.27	.28
7	Partial-Credit	.37	0.63	.44	.47
8	Partial-Credit	.36	0.53	.38	.38
9	0-1 Kodierung	.54	0.25	.23	.22
10	0-1 Kodierung	.85	0.13 min.	.15	–
11	0-1 Kodierung	.51	0.21	.28	.27
12	0-1 Kodierung	.79	0.16	.34	.36
13	Partial-Credit	.54	0.83 max.	.41	.42
14	0-1 Kodierung	.29	0.21	.40	.40
15	0-1 Kodierung	.31	0.22	.14	–
16	Partial-Credit	.87 max.	0.31	.30	.33
17	0-1 Kodierung	.74	0.19	.24	.22
18	0-1 Kodierung	.14 min.	0.12	.22	.23
19	0-1 Kodierung	.16	0.14	.44	.43
20	0-1 Kodierung	.34	0.23	.17	–
21	Partial-Credit	.75	0.57	.11	–
22	Partial-Credit	.46	0.52	.26	.24
23	0-1 Kodierung	.55	0.25	.23	.22
24	Partial-Credit	.74	0.34	.04	–
25	0-1 Kodierung	.66	0.22	.18	–
26	Partial-Credit	.56	0.62	.16	–
27	Partial-Credit	.29	0.53	.37	.38
28	Partial-Credit	.17	0.42	.54	.55
Max. 42 Punkte				Cronbachs $\alpha = .73$	Cronbachs $\alpha = .75$

Anmerkung. Fett markierte Items weisen Trennschärfen von $r_{it} < .20$ auf.

Es zeigt sich, dass die meisten Trennschärfen der verbleibenden 18 Items des Befragungsinstruments in einem eher niedrigen bis mittleren Bereich liegen. Die Itemvarianz ist bei Items, die nach dem Partial-Credit Modell kodiert wurden, erwartungsgemäß höher als bei zweistufiger Kodierung, was dafür spricht diese Kodierung beizubehalten. Die Schwierigkeitsindizes liegen insgesamt in einem zufriedenstellenden Bereich und werden in Kap. 7.2.1.3 im Zusammenhang mit dem Rasch-Modell noch genauer diskutiert. Für Skala₂₈ liegt Cronbachs α über dem akzeptablen Wert; unter Ausschluss der wenig trennscharfen Items steigt dieser Wert noch geringfügig an.

Für die Trennschärfe wurde bewusst ein sehr liberales Selektionskriterium gewählt, da das vorliegende Instrument neu entwickelt wurde und nicht auf gängige Verfahren zurückgegriffen werden konnte. Da das Hauptselektionskriterium bei den Trennschärfen lag, wurden nur Items, die ein kritisches Maß von $r_{it} = .20$ unterschritten, für weitere Skalierungsläufe entfernt. Für die nachfolgenden Analysen wurden teilweise nur noch 18 Items einbezogen, außer wenn ein Vergleich der Werte von Skala₁₈ und Skala₂₈ sinnvoll erschien.

Tabelle 16 gibt einen Überblick über die erreichten Punkte und deren Verteilung sowie die interne Konsistenz, aufgeteilt nach Substichproben. Datengrundlage für die Mittelwerte, die Standardabweichungen und die Prozentwerte der erreichten Punkte waren die Summenscores.

Tabelle 16: Mittelwert, Standardabweichung, Prozentualer Anteil erreichter Punkte, Trennschärfe und Reliabilität pro Substichproben (Skala₁₈)

Substichproben	<i>n</i>	<i>M</i>	<i>SD</i>	Erreichte Punkte %	part-whole korrigierte <i>r</i> _{it} Range	Trennschärfe <i>r</i> _{it} Durchschnitt	Cronbachs <i>α</i>
I Primar _{ohneBE}	190	10.47	3.57	40	.03 - .35 ¹	.21	.60
II Primar _{mitBE}	139	12.36	3.30	48	-.42 - .29 ²	.17	.52
III SHP	114	17.31	4.30	67	.05 - .50 ³	.34	.75

Anmerkung. Die maximale Punktzahl liegt in Skala₁₈ bei 26 Punkten. ¹Item 14 niedrigste Trennschärfe. ²Item 11 niedrigste Trennschärfe. ³Item 14 niedrigste Trennschärfe.

Die Mittelwerte steigen in den Substichproben in erwarteter Richtung. Insgesamt lösen die SHP deutlich mehr Items richtig als die Substichproben Primar_{ohneBE} und Primar_{mitBE} (genaue Analyse dazu in Kap. 7.2.5.2).

In Bezug auf die geprüften Item-Skalen „Gütekriterien“ zeigt sich, dass die Substichprobe Primar_{mitBE} sowohl bei den Trennschärfen als auch bei der Reliabilität im Vergleich zu den anderen Substichproben niedrigere Werte aufweist. Auch in den anderen Gruppen gibt es Items, die teilweise niedrige Trennschärfen zeigen. Im Falle von Substichprobe Primar_{mitBE} fällt die Trennschärfe bei Item 11 sogar negativ aus. Insgesamt liefert das Verfahren bei SHP mit und ohne Berufserfahrung die zuverlässigsten Werte, bei den übrigen Gruppen ist in Bezug auf die Genauigkeit eine vorsichtige Interpretation angezeigt.

7.2.1.2 Gütekriterien und Skalierung auf Grundlage des Rasch-Modells

In Tabelle 17 sind die Werte für den Nachweis spezifischer Objektivität zunächst für Skala₂₈ und anschließend für Skala₁₈ angegeben.

Tabelle 17: Weighted In-Fit (MNSQ) für Skala₂₈ und Skala₁₈

Item Nr.	Skala ₂₈		Skala ₁₈	
	Weighted In-Fit MNSQ	T	Weighted In-Fit MNSQ	T
1	1.16	2.8	–	–
2	1.01	0.1	1.09	1.3
3	1.11	2.2	–	–
4	1.06	0.9	–	–
5	0.94	–0.8	0.96	–0.6
6	0.98	–0.4	1.01	0.2
7	0.91	–1.9	0.96	–0.8
8	0.95	–0.9	1.01	0.3
9	1.00	–0.2	1.06	1.9
10	0.99	0.0	–	–
11	0.97	–1.3	1.03	0.8
12	0.92	–1.1	0.93	–1.1
13	0.96	–0.9	1.02	0.4
14	0.91	–1.8	0.92	–1.7
15	1.03	0.7	–	–
16	0.99	–0.4	0.98	–0.1
17	0.98	–0.3	1.03	0.5
18	1.00	0.0	1.00	0.0
19	0.90	–1.2	0.88	–1.5
20	1.03	0.7	–	–
21	1.17	2.6	–	–
22	1.03	0.7	1.12	2.2
23	1.00	0.1	1.05	1.5
24	1.12	2.1	–	–
25	1.01	0.2	–	–
26	1.12	2.4	–	–
27	0.96	–0.6	1.03	0.2
28	0.82	–2.2	0.81	–2.3

Anmerkung. Fett markierte Items weisen T-Werte außerhalb des Bereichs –2 bis +2 auf.

Die MNSQ-Werte liegen sowohl in Skala₂₈ als auch in Skala₁₈ in einem zufriedenstellenden Bereich (> 0.80 und < 1.20). In Skala₂₈ weisen die T-Werte von Item 1, 3, 21, 24, 26, 28 auf eine signifikante Abweichung einer Nullverteilung hin ($T > 2$). Mit Ausnahme von Item 28 (das die höchste Trennschärfe zeigt) wurden diese fünf Items für spätere Analysen ausgeschlossen. In Skala₁₈ liegen ebenfalls signifikante Abweichungen vor. Aufgrund des akzeptablen MNSQ-Werts und aus inhaltlichen Überlegungen wurde Item 22 im Itempool belassen. Bei Item 28 ist die Trennschärfe höher als erwartet, was in der Regel weniger problematisch ist und nicht zum Ausschluss führen muss.

Um zu prüfen, ob das Rasch-Modell auch für die einzelnen Substichproben gültig ist, wurden die MNSQ-Werte der Skala₁₈ für jede Substichprobe einzeln berechnet (Tabelle 18).

Tabelle 18: Weighted In-Fit (MNSQ) getrennt nach Substichproben (Skala₁₈)

Item Nr.	I Primar _{ohne} BE (n = 190)		II Primar _{mit} BE (n = 139)		III SHP (n = 114)	
	Weighted In-Fit MNSQ	T	Weighted In-Fit MNSQ	T	Weighted In-Fit MNSQ	T
2	1.04	0.4	0.96	-0.3	1.13	0.9
5	0.97	-0.3	1.00	0.1	0.91	-0.3
6	1.01	0.3	0.98	-0.2	1.01	0.1
7	0.98	-0.2	1.00	0.0	0.99	0.0
8	0.97	-0.3	1.04	0.5	0.98	-0.2
9	1.03	0.7	1.01	0.2	1.00	0.0
11	1.04	0.8	1.07	1.8	0.97	-0.2
12	0.94	-0.7	0.94	-0.5	0.89	-0.3
13	1.02	0.3	0.99	-0.2	1.05	0.3
14	1.01	0.2	0.98	-0.1	0.90	-1.4
16	1.00	0.0	0.97	-0.1	0.88	-0.3
17	0.96	-0.5	1.02	0.3	1.04	0.3
18	0.93	-0.2	0.99	0.0	0.97	-0.2
19	0.96	-0.1	1.01	0.2	0.95	-0.7
22	1.03	0.4	1.01	0.2	1.33	2.7
23	1.05	1.1	1.02	0.3	1.04	0.5
27	0.99	0.0	1.00	0.0	0.97	-0.2
28	0.91	-0.1	0.98	0.0	0.91	-0.9

Anmerkung: Fett markierte Items weist einen MNSQ > 1.30 auf und T-Werte außerhalb des Bereichs -2 bis +2 auf.

Die Daten zeigen, dass das Modell auch für die drei Substichproben im Einzelnen gilt und dass alle MNSQ-Werte mit Ausnahme von Item 22 in der Substichprobe SHP in einem zufriedenstellenden Bereich liegen. Item 22 weist in den anderen Substichproben allerdings gute Werte auf, weshalb das Item aus bereits genannten Gründen für weitere Analysen beibehalten wurde. Weitere Gütekriterien, die im Rahmen der Rasch-Skalierung herangezogen wurden, sind die WLE-Reliabilität, die EAP-Reliabilität sowie die Varianz der latenten Personenfähigkeiten. In Tabelle 19 werden die Werte zunächst für die gesamte Stichprobe und danach getrennt nach Substichproben für Skala₁₈ berichtet.

Tabelle 19: Gütekriterien auf Grundlage des Rasch-Modells für Skala₂₈ und Skala₁₈

Substichproben	n	Skala ₂₈			Skala ₁₈		
		Varianz	EAP-Reliabilität	WLE-Reliabilität	Varianz	EAP-Reliabilität	WLE-Reliabilität
I Primar _{ohne} BE	190	0.18	.57	.57	0.33	.62	.61
II Primar _{mit} BE	139	0.11	.51	.50	0.22	.53	.52
III SHP	114	0.42	.77	.75	0.76	.77	.74
Gesamtstichprobe	443	0.35	.74	.74	0.65	.78	.76

Insgesamt zeigt sich, dass die Werte der Skala₁₈ für alle drei Variablen deutlich zufriedenstellender sind als jene der Skala₂₈. Insbesondere ist die Varianz sowohl bezogen auf die Gesamtstichprobe als auch auf die einzelnen Substichproben auf Grundlage aller Items unbefriedigend.

Unter Ausschluss der wenig trennscharfen Items können die Varianz sowie die Reliabilitäten als ausreichend akzeptiert werden (Stancel-Piatak et al., 2013).

7.2.1.3 Itemschwierigkeitsanalyse

Mit dem Befragungsinstrument soll ein möglichst breites Wissensspektrum erfasst werden. Auskunft darüber geben der Schwierigkeitsindex (p_i) nach KTT sowie die Itemschwierigkeitsverteilung im Rahmen des Rasch-Modells, welche die Itemschwierigkeit in Abhängigkeit der Personenfähigkeit zeigt. Die Auswertungen nach KTT (Tabelle 15) hat bereits gezeigt, dass die Itemschwierigkeiten auf Grundlage der Summencores in Skala₂₈ zwischen $p_i = .14 - .87$ liegen und die Items wunschgemäß auch den obere und den unteren Wissensbereich erfassen.

In der folgenden Darstellung (Abbildung 22) werden Itemschwierigkeiten der KTT und des Rasch-Modells aufeinander bezogen und anhand der *Map of latent Distribution* dargestellt. Schwierige sowie einfache Items können mit Hilfe beider Testtheorien identifiziert werden.

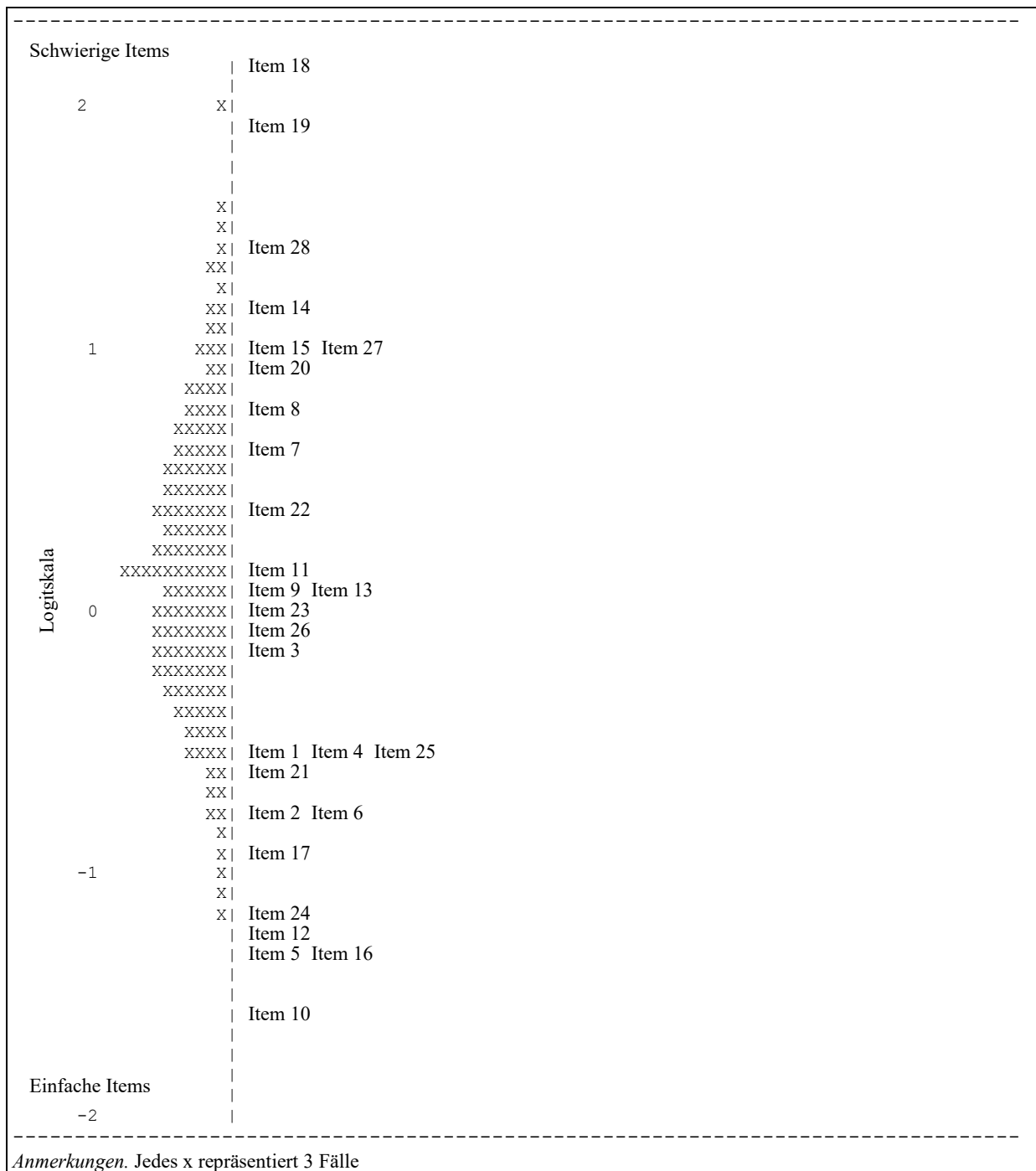


Abbildung 22. Map of latent Distribution: Gegenüberstellung der Verteilung der Itemschwierigkeit (rechts) in Relation zur Personenfähigkeit (links) im Rasch-Modell in der Skala₂₈

Es zeigt sich, dass die Items 18 und 19 in den oberen Extreimbereichen als sehr schwierig einzustufen sind. Diese Items werden von vielen Personen nur mit einer geringen Wahrscheinlichkeit gelöst. Andererseits liegen im niedrigen Wissensbereich Items vor (Item 10, 5, 16, 12), die gemäß Rasch-Modell praktisch von allen gelöst werden können. Bei den erwähnten Items gibt es also einen gewissen Boden- und Deckeneffekt, und die Verteilung ist in den Randbereichen nicht optimal.

Nach Ausschluss der wenig trennscharfen Items zeigt sich für Skala₁₈ eine bessere Passung zwischen Itemschwierigkeit und Personenfähigkeit (Abbildung 23).

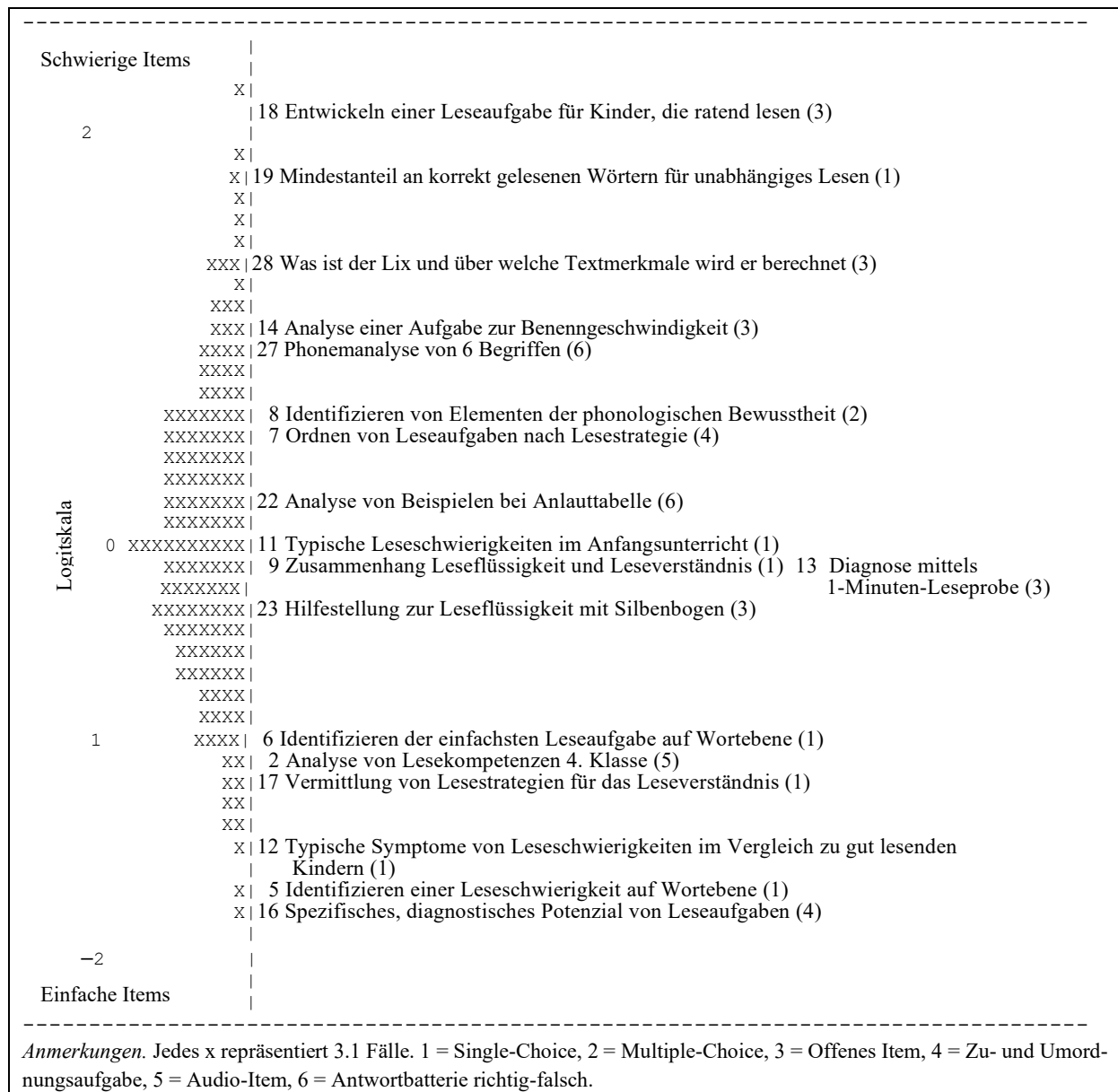


Abbildung 23. Map of latent Distribution: Verteilung der Itemschwierigkeiten (rechts) in Relation zur Verteilung der Personenfähigkeit (links) in der Hauptuntersuchung Skala₁₈ und Berücksichtigung der Itemformate

Die Streuung der Personenfähigkeit wird mit Hilfe der Skala₁₈ weitgehend erfasst. Dabei zeigt sich, dass offene Items erwartungsgemäß schwieriger sind als gebundene Formate.

In der Abbildung 24 werden die Itemschwierigkeiten der drei Substichproben verglichen. In *Conquest* werden dazu die Itemschwierigkeitsparameter der gemeinsamen Skala aller drei Stichproben für die separate Schätzung der Personenparameter in den Substichproben importiert und „verankert“. Damit wird eine gemeinsame Skala fixiert, und es lässt sich darstellen, wie sich die Verteilung der Personenfähigkeit auf dieser Skala verschiebt.

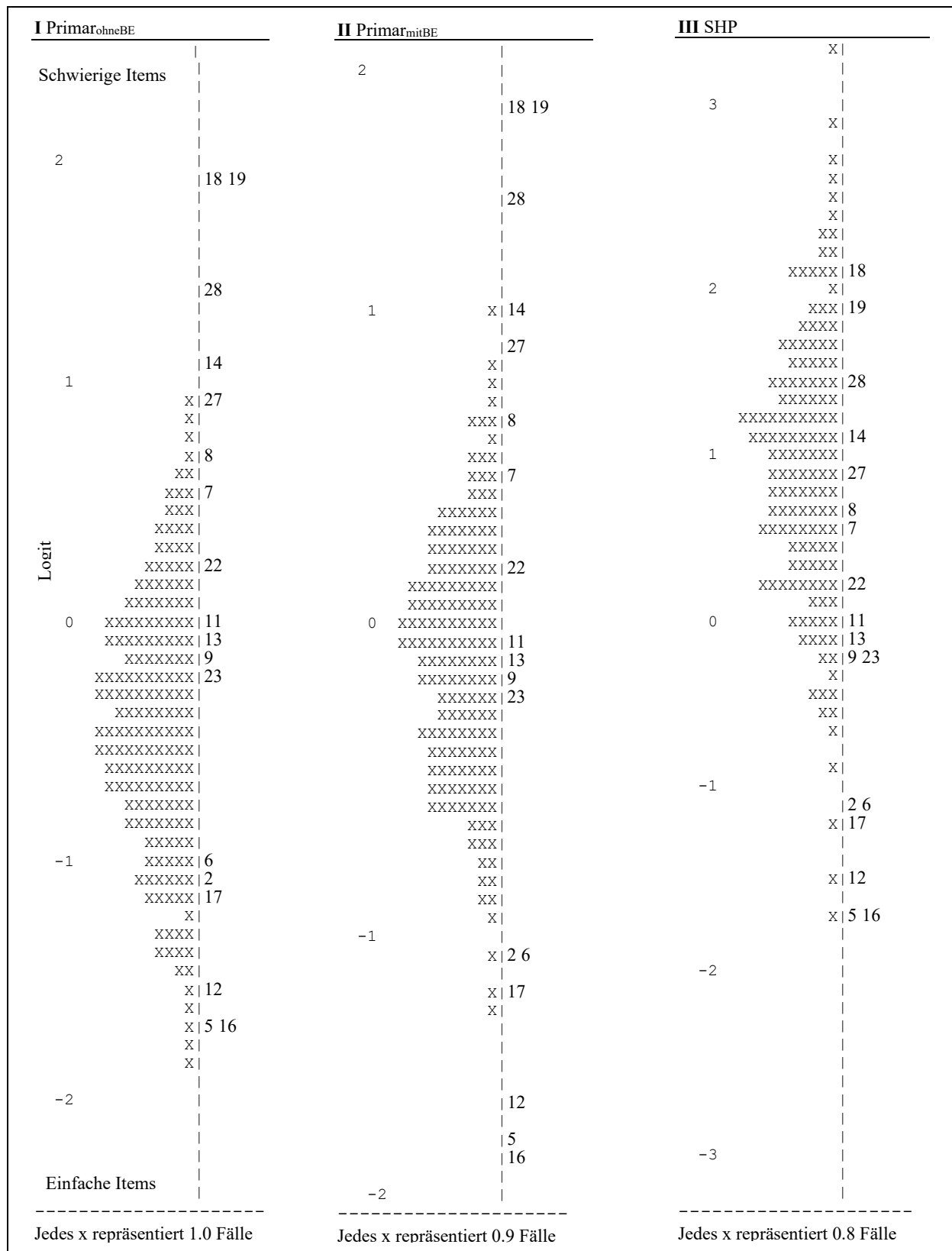


Abbildung 24. Itemschwierigkeit der Substichprobe I, II und III im Vergleich bei gemeinsamer fixierter Skala (Skala₁₈).

In Abbildung 24 zeigt sich, wie sich die Verteilung der Personen der Substichproben I bis III nach oben verschiebt. Dies gilt bei fixierter gemeinsam geschätzter Itemschwierigkeit unter Einbezug aller drei Substichproben. Es wird sichtbar, dass die Lösungswahrscheinlichkeit z.B. für Item 18 und Item 19 für einen Großteil der Studierenden der Substichproben Primar_{ohne}BE

und $\text{Primar}_{\text{mitBE}}$ sehr gering ist, während diese Items von Studierenden der Substichprobe SHP mit hohen Fähigkeiten noch lösbar sind. Weiter ergibt sich, dass das Fähigkeitsspektrum der Substichprobe SHP durch die Items am besten abgedeckt ist, was sich bereits aus der Varianz von 0.76 (Tabelle 19) schließen lässt. Dagegen vermögen die Items das Fähigkeitsspektrum von Substichprobe $\text{Primar}_{\text{mitBE}}$ nur ungenügend abzubilden, was ebenfalls bereits aus der ermittelten Varianz von 0.22 hervorgeht.

Beispiel-Items

Punkto Inhaltes lässt sich feststellen, dass Items mit offenem Antwortformat, bei denen die Teilnehmenden selbst Aufgaben entwickeln (Item 18) oder über spezifische Fachbegriffe wie etwa den Lix Auskunft geben müssen (Item 28), ein großes Wissen erfordern.

Folgendes Beispiel zeigt eine typische Antwort zu Item 18, die das Vorliegen eines vertieften fachdidaktischen Verständnisses zum Leseprozess erfordert (Abbildung 25).

Kinder mit Leseschwächen kompensieren ihr Defizit manchmal, indem sie **ratend** lesen. Das Leseverständnis ist dadurch beeinträchtigt.

Mit welchen Leseaufgaben finden Sie als Lehrperson heraus, ob das Kind wirklich genau und sinnverstehend liest?

Formulieren Sie **eine konkrete Leseaufgabe** für das Kind, die zum **genauen und sinnverstehenden** Lesen auf **Satzebene** zwingt. Das Kind soll dabei **nicht** laut vorlesen.

Falls Sie die Antwort(en) nicht wissen, schreiben Sie „ich weiss es nicht“.

Stille Leseaufgabe für das Kind (umschreiben Sie die Aufgabe nicht, sondern stellen Sie sie direkt)

Marco geht auf den Spielplatz nachdem er schwimmen war.

Was macht Marco zuerst? ☐ Er geht auf den Spielplatz.
☐ Er geht schwimmen.

Abbildung 25: Beispiel-Item18 mit hoher Wissensanforderungen (offenes Antwortformat)

Neben offenen Antwortformaten gibt es auch schwierige Items mit gebundenem Antwortformat. Item 19 kann hier als Beispiel dienen (Abbildung 26).

Um einen Text **richtig** und **ohne Hilfe** verstehen zu können, muss ein bestimmter Anteil an Wörtern korrekt gelesen werden. Wie hoch muss dieser Anteil **mindestens** sein? (Eine Antwort ankreuzen)

a	50%-60%	<input type="checkbox"/>
b	60%-70%	<input type="checkbox"/>
c	70%-80%	<input type="checkbox"/>
d	80%-90%	<input type="checkbox"/>
e	90%-100%	<input type="checkbox"/>

Abbildung 26: Beispiel-Item 19 mit hoher Wissensanforderungen (gebundenes Antwortformat)

Wird das Antwortverhalten der untersuchten Subgruppen (inklusive Laien) für Item 19 getrennt betrachtet, so zeigt sich folgendes Bild (Tabelle 20).

Tabelle 20: Antwortverhalten Beispiel-Item 19 getrennt nach Substichproben

Distraktoren	0 Laien %	I Primar _{ohneBE} %	II Primar _{mitBE} %	III SHP %	Gesamtstichprobe %
a	7.8	2.6	2.9	4.4	3.7
b	26.6	23.7	22.3	5.3	19.5
c	48.4	43.2	47.5	14.9	38.7
d	15.6	24.2	24.5	26.3	23.7
e ^a	1.6	6.3	2.9	49.1	14.4

Anmerkung. % = Angabe des prozentualen Anteils angekreuzter Distraktoren. ^aAntwort e ist korrekt.

Insgesamt haben nur 14.4% der Gesamtstichprobe die Antwort e als richtige Lösung erkannt. Am plausibelsten war für die Befragten Antwort c, die von 38.7% gewählt wurde. Von Laien ist Item 19 kaum zu lösen, wohingegen es von fast 50.0% der SHP richtig beantwortet wurde. Substichprobe Primar_{mitBE} schnitt mit nur 2.9% richtigen Lösungen unerwartet schlecht ab. Diese Ergebnisse weisen darauf hin, dass die Inhalte der vier schwierigsten Items im Rahmen der Ausbildung in Schulischer Heilpädagogik vermittelt werden, dass sie jedoch während der Grundausbildung möglicherweise nicht zum Studienstoff gehören. Es scheint sich hier entsprechend um ausbildungsspezifisches Wissen zu handeln. Diese Vermutung wird im Rahmen der differenziellen Itemfunktionsanalyse später in Kap. 7.2.3.3 noch diskutiert.

Exemplarisch für ein besonders einfaches Item steht Item 5 (Abbildung 27). Es ist bewusst einfach konzipiert und wurde im Sinne einer „warm-up“-Frage am Anfang des Befragungsinstruments platziert.

Ein Kind am Ende der 2. Klasse liest in zügigem Tempo < r' o' s' a' > (jeden Buchstaben einzeln). Womit hat dieses Kind Mühe? (Eine Antwort ankreuzen)		
Das Kind hat Mühe...		
a	mit der auditiven Unterscheidung der Laute	<input type="checkbox"/>
b	mit dem Lautieren der Buchstaben	<input type="checkbox"/>
c	mit dem Zusammenlauten von Buchstaben	<input type="checkbox"/>
d	mit dem auditiven Arbeitsgedächtnis	<input type="checkbox"/>
e	mit der phonologischen Bewusstheit	<input type="checkbox"/>

Abbildung 27. Beispiel-Item 5 mit niedriger Wissensanforderungen (gebundenes Antwortformat)

Wird das Antwortverhalten der untersuchten Substichproben für Item 5 (inklusive Laien) getrennt betrachtet, so zeigt sich folgendes Bild (Tabelle 21).

Tabelle 21: Antwortverhalten Beispiel-Item 5 getrennt nach Substichproben

Distraktoren	0 Laien %	I Primar _{ohneBE} %	II Primar _{mitBE} %	III SHP %	Gesamtstichprobe %
a	0	0	0.7	0	0.1
b	0	1.7	1.4	0	1.0
c ^a	76.6	71.7	87.0	93.0	81.3
d	4.7	1.7	2.9	4.1	3.1
e	18.7	24.9	8.0	2.9	14.5

Anmerkung. % = Angabe des prozentualen Anteils angekreuzter Distraktoren. ^aAntwort c ist korrekt.

Insgesamt lässt sich sagen, dass 81.3% der Teilnehmenden die Antwort **c** als richtige Antwort identifiziert haben. Distraktor **a** erschien hingegen nicht plausibel und wurde von kaum jemandem gewählt. Bemerkenswert scheint zudem, dass Antwort **c** auch für Laien plausibel war und dass das Item wohl auch intuitiv aus den vorgegebenen Antworten abgeleitet oder erraten werden konnte. Allerdings wurde von den Laien wie auch von Substichprobe Primar_{ohne BE} Antwort **e** relativ häufig angekreuzt, was zeigt, dass das Konzept der phonologischen Bewusstheit in diesen Substichproben nicht oder falsch verstanden war. Auch hier zeigt sich, dass die Substichprobe SHP mit 93.0% richtigen Lösungen am besten abschneidet.

7.2.1.4 Fazit aus der Item-Skalenanalyse

Von den ursprünglich 28 Items weisen 18 Items akzeptable Trennschärfen auf. Unter Ausschluss der kritischen Items können dem Befragungsinstrument akzeptable Itemvarianzen und eine ausreichende Reliabilität im Sinne interner Konsistenz gemäß KTT attestiert werden. Die Werte auf der Grundlage des Rasch-Modells sind für die Gesamtstichprobe etwas günstiger, und aufgrund dieser Analysen müssten lediglich sechs Items ausgeschlossen werden. Allerdings scheint es sinnvoll, das trennschärfenbasierte Selektions-Kriterium auch vor dem Hintergrund des Rasch-Modells zu berücksichtigen, da hier theoretisch von gleichen Trennschärfen für alle Items ausgegangen wird.

Betrachtet man die Werte der Item-Skalenanalyse (Skala₁₈) in den einzelnen Substichproben getrennt, zeigen sich nach KTT für die Substichprobe SHP zufriedenstellende Werte (Trennschärfe, Itemvarianz, Interne Konsistenz). Auch die Skalenwerte auf der Grundlage des Rasch-Modells (Varianz, EAP- und WLE-Reliabilität) sind für diese Substichprobe ausreichend bis gut. Die Eignung des Befragungsinstruments muss für die Substichprobe Primar_{mitBE} aufgrund problematischer Testkennwerte allerdings in Frage gestellt werden.

Im Weiteren kann dem Befragungsinstrument Skala₁₈ insgesamt unter Einbezug der gesamten Stichprobe eine angemessene Schwierigkeit bescheinigt werden. Insbesondere kann das Instrument in den oberen und den unteren Leistungsbereichen gut differenzieren. Durch die Kodierung mit dem Partial-Credit-Modell bei insgesamt 14 Items kann zudem eine differenzierte Verteilung der Schwierigkeiten erzielt werden. Im mittleren Schwierigkeitsbereich wären noch weitere Items erstrebenswert. Mit Blick auf die Schwierigkeitsverteilung in den einzelnen Substichproben zeigt sich wiederum, dass das Fähigkeitsspektrum der Substichprobe SHP am besten abgedeckt ist. In der Substichprobe Primar_{ohneBE} gelingt dies nur bedingt, und eher ungünstig zeigt sich die Verteilung bei der Substichprobe Primar_{mitBE}.

7.2.2 Ergebnisse der Distraktorenanalyse

Die Distraktorenanalyse wurde für alle Single- und Multiple-Choice-Aufgaben sowie für die Audio-Items, die ebenfalls nach dem Muster von Multiple-Choice-Aufgaben konstruiert sind, durchgeführt. Bei allen Aufgaben war jeweils angegeben, ob *eine* oder *mehrere* Antworten anzukreuzen sind.

Die Distraktoren von 21 Items wurden hinsichtlich ihrer Plausibilität beurteilt. Insgesamt enthält das ganze Instrument 96 Distraktoren. Die Analyse erfolgte zunächst für die Substichproben I, II und III und anschließend nur für die Laienstichprobe. Die folgende Übersicht (Tabelle 22) zeigt, welche Distraktoren über die Stichproben hinweg von weniger als 5% der Studienteilnehmenden angekreuzt wurden. Werden Distraktoren sowohl von Professionellen (Substichproben I, II, III) als auch von Laien kaum gewählt, so liegt der Schluss nahe, dass es sich um ungünstig formulierte Distraktoren handelt, die nicht als echte Ablenker funktionieren.

Tabelle 22: Analyse zur Beurteilung der Plausibilität von Distraktoren

Item Nr.	Substichproben I II III	Substichprobe 0 Laien	Distraktoren über alle Substichproben hinweg unplausibel
	Distraktor von < 5% der Teilnehmenden gewählt		
1 ^a	– e Kategorie „übertroffen“ wurde von 1.1% angekreuzt	d e Kategorie „übertroffen“ wurde nicht angekreuzt	1 e
2	– f Kategorie „übertroffen“ wurde nicht angekreuzt	e f Kategorie „übertroffen“ wurde nicht angekreuzt	2 f
3 ^a	f	f	3 f
4	Alle Distraktoren plausibel		
5	a b d	a b d	5 a b d
6	a c e	–	
7	Alle Distraktoren plausibel		
8	f	–	
9	Alle Distraktoren plausibel		
10 ^a	a c e	a c e	10 a c e
11	d	d	11 d
12	c d	c	12 c
15	a	–	
16	d	–	
17	d	–	
19	a	e	
20	Alle Distraktoren plausibel		
21	Alle Distraktoren plausibel		
22	Alle Distraktoren plausibel		
24 ^a	d	d	24 d
25	Alle Distraktoren plausibel		

Anmerkung. Fett markierte Distraktoren wurden sowohl von der Grundstichprobe als auch von Laien als unplausibel erkannt.

^a Items wurden aufgrund niedriger Trennschärfen im finalen Befragungsinstrument ausgeschlossen

Unter Anwendung der 5%-Regel fallen 21 Distraktoren auf, die für die Studienteilnehmenden (Substichproben I, II, III) offensichtlich zu unattraktiv waren und kaum gewählt wurden. In einigen Fällen waren die Distraktoren selbst für Laien nicht plausibel, so dass Formulierungsschwächen vorliegen mussten, die das Antwortverhalten beeinflusste. Wurden Distraktoren von

Laien aber gewählt, so leisteten diese zumindest einen Beitrag zur Validität des Instruments, da berufsgruppenspezifisches Wissen abgefragt wurde, das in der Laienstichprobe nicht vorlag. Von den im definitiven Befragungsinstrument verbleibenden Items waren in diesem Sinne nur Distraktor **2f**, die Distraktoren **5 abd**, Distraktor **11d** und Distraktor **12c** problematisch. Für die weitere Verwendung des Befragungsinstruments müssen diese überarbeitet werden oder können im Fall von Item 2, 11 und 12 sogar weggelassen werden, da sie keinen Informationsgewinn bringen.

Bei Item 1 und 2 handelte es sich um Audio-Items, bei denen in beiden Fällen Kinder mit deutlichen Leseschwierigkeiten aufgrund von Hörbeispielen beurteilt werden mussten (Abbildung 28). Die Antwort „noch nicht erreicht“ war hier korrekt. Die Studienteilnehmenden erkannten durchwegs, dass die Ziele für das Lesen am Ende der ersten Klasse (Item 1) bzw. der vierten Klasse (Item 2) nicht „übertroffen“ wurden.

Das Kind steht am Ende der 4. Klasse . Kreuzen Sie alle Kompetenzen an, die das Kind mehrheitlich zeigt. (Mehrere Antworten ankreuzen) Nehmen Sie zum Schluss eine Gesamteinschätzung vor (f). (Eine Antwort ankreuzen)		
Alter: 9 Jahre 10 Monate Stufe: Ende 4. Klasse. (Juli)		
a	Lesefehler und Auslassungen werden mehrheitlich selbst korrigiert.	<input type="checkbox"/>
b	Größere Einheiten (z.B. Silben) werden mehrheitlich genutzt.	<input type="checkbox"/>
c	Sichtwortschatz bei kurzen Wörtern (bis 5 Buchstaben) mehrheitlich vorhanden.	<input type="checkbox"/>
d	Sichtwortschatz bei längeren/komplexeren Wörtern (mehr als 5 Buchstaben) mehrheitlich vorhanden.	<input type="checkbox"/>
e	Ausdrucksvolle Betonung / Prosodie mehrheitlich angemessen.	<input type="checkbox"/>
f	Gesamteinschätzung Die Ziele für das Lesen am Ende der 4. Klasse sind...	übertroffen <input type="checkbox"/> erreicht <input type="checkbox"/> noch nicht erreicht <input type="checkbox"/>

Abbildung 28. Beispiel-Item 2 (Audio-Item) mit unplausibler Kategorisierung bei Distraktor f

Die Kategorie „übertroffen“ ist bei Distraktor **f** unplausibel, weshalb er für die weitere Verwendung eliminiert werden kann. Selbst Laien erkannten, dass die zu erwartenden Leseleistungen der Kinder in einem normalen oder niedrigen Bereich liegen und sicher *nicht* „übertroffen“ wurden. Alle Studierenden der Substichprobe SHP mit Zusatzausbildung in Schulischer Heilpädagogik schätzten das Kind korrekt ein (*Die Ziele für das Lesen am Ende der 4. Klasse sind noch nicht erreicht*) und erkannten die Leseschwierigkeiten. Bei $\text{Primar}_{\text{mitBE}}$ wurden die Leseschwierigkeiten von 92.8%, bei $\text{Primar}_{\text{ohneBE}}$ von 88.4% der Probandinnen und Probanden erkannt.

Item 5 ist insofern auffallend, als die „Unplausibilität“ gleich für mehrere Distraktoren – und zudem in identischer Weise auch für die Laienstichprobe – zutrifft (Abbildung 29).

Ein Kind am Ende der 2. Klasse liest in zügigem Tempo < r'o's'a' > (jeden Buchstaben einzeln). Womit hat dieses Kind Mühe? (Eine Antwort ankreuzen)		
Das Kind hat Mühe...		
a	mit der auditiven Unterscheidung der Laute	<input type="checkbox"/> Nicht plausibel
b	mit dem Lautieren der Buchstaben	<input type="checkbox"/> Nicht plausibel
c	mit dem Zusammenlauten von Buchstaben	<input type="checkbox"/>
d	mit dem auditiven Arbeitsgedächtnis	<input type="checkbox"/> Nicht plausibel
e	mit der phonologischen Bewusstheit	<input type="checkbox"/>

Abbildung 29. Beispiel-Item 5 mit unplausiblen Distraktoren a, b, d

Da Item 5 als „warum-up“-Frage konzipiert ist und eine zufriedenstellende Trennschärfe aufweist, wurde es in Skala₁₈ dennoch für weitere Berechnungen im Itempool belassen. Für die weitere Verwendung des Instrumentes ist aber zu überlegen, welche geeigneteren Distraktoren formuliert werden können.

7.2.3 Ergebnisse der differenziellen Itemfunktionsanalyse (DIF)

Eine Voraussetzung für die Analyse differenzieller Itemfunktionen ist die Gültigkeit des Rasch-Modells für die gesamte Stichprobe (Blömeke et al., 2009). Die Gültigkeit des eindimensionalen Rasch-Modells konnte anhand der Fit-Statistiken bestätigt werden.

Zur Klärung der Frage, ob Aufgaben in den verschiedenen Teilstichproben dasselbe Konstrukt messen und damit Messinvarianz vorliegt, wurden die drei Substichproben, die nach Ausbildungsniveau gebildet wurden, verglichen. Zur Analyse der differenziellen Itemfunktion (DIF-Analyse, Kap. 6.3.3) wurden nur die 18 im Itempool verbleibenden Items einbezogen. Datengrundlage für den Vergleich waren die Schwierigkeitsparameter (Logits), die so festgelegt wurden, dass der Mittelwert der Itemschwierigkeiten pro Gruppe je Null war. In *ConQuest* wurde dafür die Einstellung („constraints = Items“) ausgeführt. Dies hat dazu geführt, dass im Folgenden immer die Abweichungen vom Ideal der Nulldifferenz der Logit-Werte interpretiert werden.

Zur Veranschaulichung werden im Folgenden zunächst die Abweichungen der Schwierigkeitsparameter graphisch dargestellt. Anschließend werden die Abweichungen statistisch abgesichert.

7.2.3.1 Graphische Inspektion der differenziellen Itemfunktion

In der folgenden Graphik (Abbildung 30) sind die Itemschwierigkeitsparameter der drei Substichproben der Skala₁₈ dargestellt und für jede Gruppe so normiert, dass ihr Mittelwert pro Gruppe Null ist. Daraus lassen sich die *relativen* Abweichungen der drei Substichproben ablesen. Bei Items, wo die Punkte der Substichproben weit voneinander entfernt liegen, muss davon

ausgegangen werden, dass diese stichprobenabhängig messen und dass somit Messvarianz vorliegt.

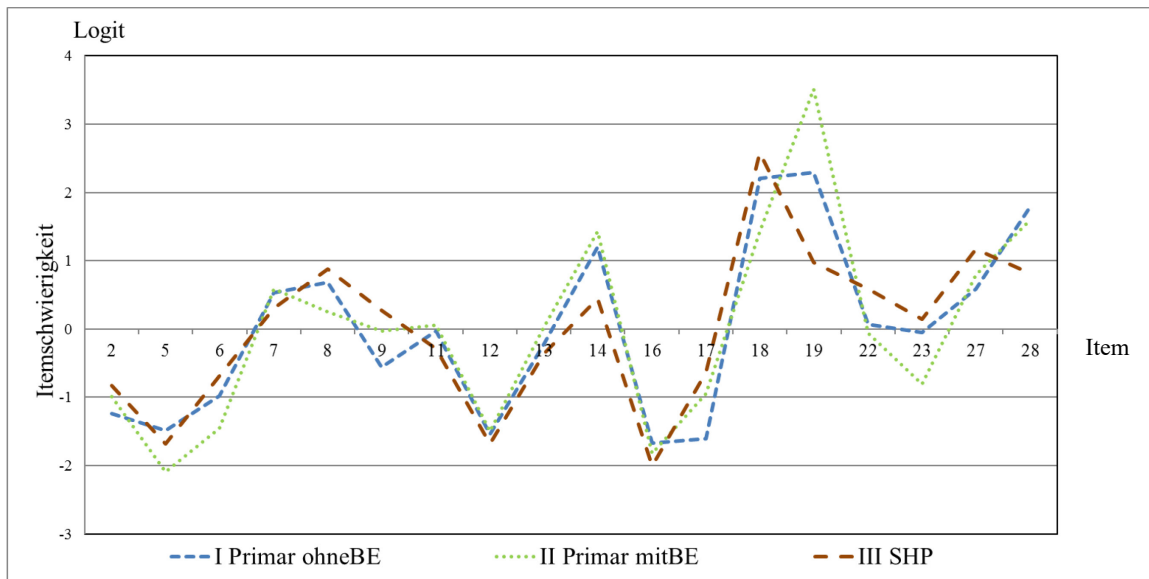


Abbildung 30. Vergleich der drei Substichproben hinsichtlich ihrer relativen Abweichung der Itemschwierigkeit vom Mittelwert Null in Skala₁₈

In der Graphik wird ersichtlich, dass einige Items sehr unterschiedlich beantwortet wurden und dass von einer DIF ausgegangen werden muss. Sehr augenfällig ist dies etwa bei Item 9, 14, 17, 19, 23 und 28. Auffallend ist, dass die DIF über alle drei Substichproben festgestellt werden muss und sich die Disparallelität bei verschiedenen Items bei jeweils anderen Subgruppen zeigt. So wurde z.B. Item 14 von den beiden Substichproben Primar_{ohneBE} und Primar_{mitBE} ähnlicher gelöst als von der Substichprobe SHP. Anders sieht es z.B. bei Item 23 aus, wo die Substichproben Primar_{ohneBE} und SHP ähnliche Schwierigkeitsparameter aufweisen, wohingegen das Item bei Substichprobe Primar_{mitBE} andere Eigenschaften zu messen scheint.

Insbesondere zwischen den Substichproben Primar_{mitBE} und SHP muss die postulierte Stichprobenunabhängigkeit hinterfragt werden, so dass hier nochmals ein auf die beiden Substichproben fokussierter graphischer Modelltest vorgenommen wird. Das folgende bivariate Streudiagramm (Abbildung 31) zeigt die relativen Abweichungen der Itemschwierigkeit, diesmal aber im paarweisen exemplarischen Vergleich zwischen den Substichproben Primar_{mitBE} und SHP. Je stärker die Abweichung von der Diagonalen, desto größer die DIF des jeweiligen Items. Umgekehrt gilt: Je genauer die Itemschwierigkeitsparameter auf der Diagonalen liegen, desto besser passt das Modell gleichzeitig für *beide* Substichproben und desto eindeutiger ist von generalisierter Rasch-Homogenität auszugehen (Moosbrugger & Kelava, 2012b).

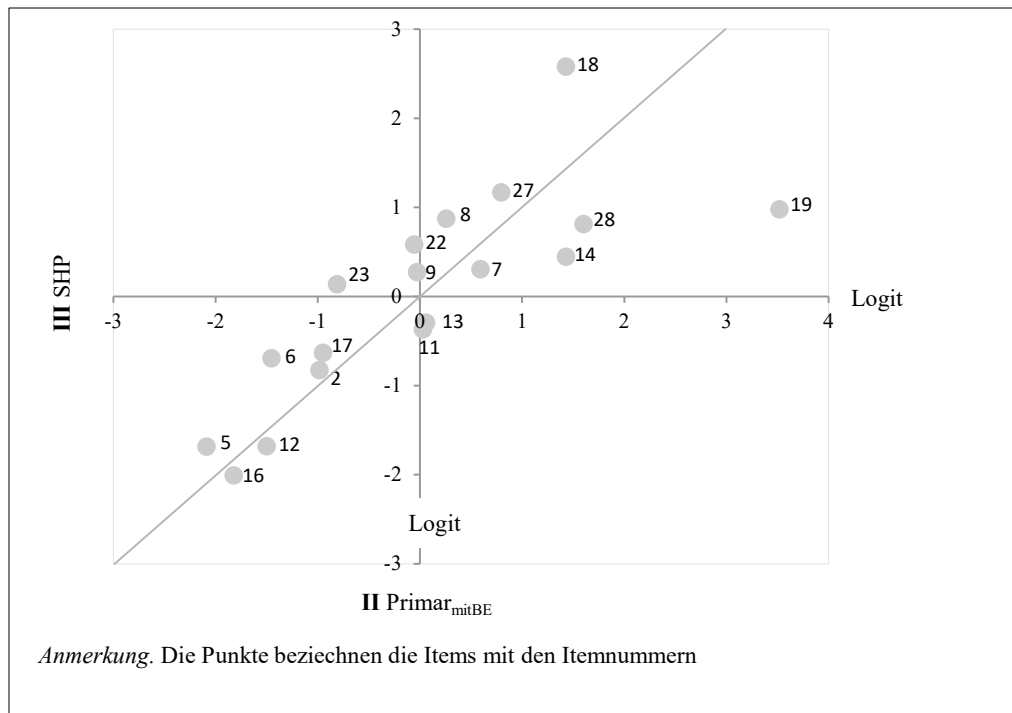


Abbildung 31. Streudiagramm zum Vergleich der relativen Itemschwierigkeit von Substichprobe PrimarohneBE und Substichprobe SHP

Im bivariaten Streudiagramm zeigt sich nochmals, bei welchen Items die größten Abweichungen der Logits zu verzeichnen sind. Besonders deutliche Unterschiede zeichnen sich bei den Items 18 und 19, aber auch bei den Items 6, 14, 23 und 28 ab. Ein inhaltlicher Erklärungsversuch wäre, dass Studierende mit einer Zusatzausbildung in Schulischer Heilpädagogik auf andere Konzepte zurückgreifen können als Studierenden am Ende der Grundausbildung. Item-Beispiele dazu folgen im Anschluss an die statistischen Auswertungen zum Vorliegen von DIF.

7.2.3.2 Statistische Analyse der differenziellen Itemfunktion

In der folgenden Zusammenschau (Tabelle 23) werden die Itemschwierigkeiten (Betrag der Differenz der Logit-Werte) der drei Substichproben paarweise verglichen und Abweichungen von der Nulldifferenz statistisch abgesichert (Kap. 6.3.3). Die zugrundeliegenden paarweisen Auswertungen mit den entsprechenden, für die Berechnung notwendigen Werte (Logit und Standardfehler, Signifikanz), finden sich im Anhang (Tabelle 35, Tabelle 36, Tabelle 37).

Tabelle 23: Differenzielle Itemfunktion in den drei Substichproben im paarweisen Vergleich

Item Nr.	Substichprobe I vs. II		Substichprobe II vs. III		Substichprobe I vs. III	
	Betrag Differenz der Logit-Werte	<i>t</i>	Betrag Differenz der Logit-Werte	<i>t</i>	Betrag Differenz der Logit-Werte	<i>t</i>
2	0.26	1.37	0.17	0.60	0.43	1.57
5	0.60 <i>moderat</i>	2.04*	0.41	0.99	0.20	0.54
6	0.48 <i>moderat</i>	1.89	0.76 <i>hoch</i>	2.40*	0.28	0.99
7	0.06	0.31	0.28	1.42	0.23	1.15
8	0.43 <i>moderat</i>	2.28*	0.62 <i>moderat</i>	3.09**	0.17	0.90
9	0.53 <i>moderat</i>	2.32*	0.31	1.14	0.83 <i>hoch</i>	3.28**
11	0.09	0.39	0.35	1.26	0.26	0.97
12	0.04	0.16	0.18	0.46	0.14	0.37
13	0.25	1.72	0.39	1.95	0.14	0.72
14	0.23	0.81	0.98 <i>hoch</i>	3.31**	0.75 <i>hoch</i>	2.69**
16	0.15	0.59	0.19	0.34	0.34	0.66
17	0.65 <i>hoch</i>	2.62**	0.32	1.05	0.97 <i>hoch</i>	3.32**
18	0.79 <i>hoch</i>	2.26*	1.15 <i>hoch</i>	3.50**	0.37	1.99
19	1.22 <i>hoch</i>	2.19*	2.55 <i>hoch</i>	4.87**	1.31 <i>hoch</i>	3.87**
22	0.12	0.71	0.64 <i>moderat</i>	3.09**	0.52 <i>moderat</i>	2.63*
23	0.76 <i>hoch</i>	3.22**	0.95 <i>hoch</i>	3.44**	0.19	0.75
27	0.20	1.09	0.38	1.80	0.58 <i>moderat</i>	2.95**
28	0.21	0.57	0.78 <i>hoch</i>	3.03**	0.99 <i>hoch</i>	3.19**

Anmerkung. Grau hinterlegte Items weisen bei keinem Vergleich DIF auf. * $p < .05$ ** $p < .01$.

Die Zusammenschau der abweichenden Logit-Beträge und T-Werte zeigt, dass über die drei Substichproben hinweg lediglich sechs Items Messinvarianz aufweisen. Weitere vier Items (5, 8, 22, 27) weisen jeweils bei mindestens einem Vergleich signifikante Differenzen der Logit-Beiträge auf. Die Abweichung liegen aber unter dem kritischen Wert von 0.64 und sind daher als „moderat“ einzustufen (Tristán, 2006). Die übrigen acht Items (Abweichung „hoch“) werden von den drei Substichproben deutlich unterschiedlich gelöst und damit in den Substichproben jeweils andere Wissensaspekte gemessen. Am deutlichsten sind die Unterschiede im Vergleich der Substichprobe Primar_{mitBE} und SHP sowie der Substichprobe Primar_{ohneBE} und SHP, wo die meisten Abweichungen vorliegen. Item 19 funktioniert über alle drei Substichproben hinweg unterschiedlich.

Bei fehlender Messinvarianz würden die entsprechenden Items normalerweise aus dem Itempool entfernt. Im vorliegenden Fall würde dies allerdings zu einer massiven Einschränkung der Validität führen. Um dies zu verhindern, werden deshalb zwei pragmatische Strategien verfolgt (Kap. 6.3.3). Zum einen geben die Differenzen Anlass zu einer inhaltlichen Interpretation, die den Bias erklären könnte, und zum anderen werden Items „gesplittet“ und verschiedene Modelle, in denen Messinvarianz operationalisiert wurde, hinsichtlich ihrer Robustheit auf das Gesamtergebnis überprüft.

7.2.3.3 Interpretation der differenziellen Itemfunktion

Zunächst folgt ein inhaltlicher Erklärungsversuch zur DIF, der an ausgewählten Items exemplarisch erläutert wird. Dieser interpretativen Strategie liegt die Idee zugrunde, DIF nicht einfach als Bias im Sinne einer Benachteiligung und Unfairness gegenüber einer Gruppe zu betrachten. Vielmehr könnte sie auch als Indiz für unterschiedliche Lern- und Lösungsprozesse, angewendete Konzepte und Erfahrungen interpretiert werden. Allerdings lassen sich nicht alle Differenzen sinnvoll interpretieren. Zudem müssen die Interpretationen mit aller Vorsicht gelesen werden, da es sich um nicht überprüfte Vermutungen handelt. Im Folgenden werden Items dargestellt, bei denen deutliche Lösungsmuster auftreten und *hohe* Abweichungen bestehen. Items in Gruppenvergleichen mit *moderaten* Abweichungen werden der Vollständigkeit halber genannt, aber nicht weiter analysiert.

Items zu spezifisch heilpädagogischem Wissen

Item 14, 22 und 28 werden von den Substichproben $\text{Primar}_{\text{ohneBE}}$ und $\text{Primar}_{\text{mitBE}}$ gleich gut gelöst. Unterschiede zeigen sich aber im Vergleich zur Substichprobe SHP. Die Unterschiede zur Substichprobe SHP können dahingehend erklärt werden, dass bei gewissen Items Konzepte gefragt sind, die erst im Rahmen der Zusatzausbildung erworben werden und auf die zur Lösung von Items zurückgegriffen werden kann. Bei den Items 14 und 28 handelt es sich um den Aspekt der „Benennungsgeschwindigkeit“ und den „Lesbarkeitsindex“ (Lix). Item 22 (Abbildung 32) ist zudem für die Substichprobe SHP verhältnismäßig zu schwierig. Hier haben sich SHP teilweise von einer Fehlvorstellung leiten lassen und bei Distraktor **a** häufiger die falsche Antwort gegeben.

Kinder lernen die Buchstaben oft mit Hilfe von Anlauttabellen. Bei folgenden Vergleichen gibt es immer eine bessere und eine weniger geeignete Variante, um die Buchstaben zu erarbeiten. Kreuzen Sie jeweils an, ob die Aussage <i>stimmt</i> oder <i>nicht stimmt</i> .		stimmt	stimmt nicht
a	Es ist besser den Buchstaben < B > mit Besen als mit Baum zu erarbeiten.	<input type="checkbox"/>	<input type="checkbox"/>
b	Es ist besser den Buchstaben < T > mit Tintenfisch als mit Tiger zu erarbeiten.	<input type="checkbox"/>	<input type="checkbox"/>
c	Es ist besser den Buchstaben < F > mit Feder als mit Frosch zu erarbeiten.	<input type="checkbox"/>	<input type="checkbox"/>
d	Es ist besser den Buchstaben < Z > mit Zeit als mit Zahn zu erarbeiten.	<input type="checkbox"/>	<input type="checkbox"/>

Abbildung 32. Beispiel-Item 22 zur Analyse von Anlauttabellen

Items mit inhaltlicher Nähe zu Berufserfahrungswissen

Bei den Items 6, 8, 18 und 23 fällt auf, dass sie in den Substichproben $\text{Primar}_{\text{ohneBE}}$ und SHP gleich funktionieren, nicht aber in der Substichprobe $\text{Primar}_{\text{mitBE}}$. Die Items sind für $\text{Primar}_{\text{mitBE}}$ einfacher, als dies im Mittel zu erwarten gewesen wäre. Hier könnte die Berufserfahrung der $\text{Primar}_{\text{mitBE}}$ einen Einfluss auf das Lösungsverhalten haben. Die DIF sagt hier aber lediglich aus, dass von der Substichprobe $\text{Primar}_{\text{mitBE}}$ ein *anderes* Wissen/Können aktiviert wurde als von den übrigen Substichproben, und dass das Item im Vergleich überproportional viel einfacher war – nicht aber absolut.

Welche unterschiedlichen Prozesse hier möglicherweise ablaufen, kann exemplarisch an Item 6 gezeigt werden (Abbildung 33).

Ein Kind soll zu Beginn der 1. Klasse folgende Aufgaben lösen. Welches ist die einfachste Aufgabe ? (Eine Antwort ankreuzen)		
a	Was wird aus < Ball > ohne /b/ ?	<input type="checkbox"/>
b	Hörst du ein /s/ im Wort < Musik > ?	<input type="checkbox"/>
c	Welche Laute hörst du im Wort < Ente > ?	<input type="checkbox"/>
d	Hörst du ein /n/ im Wort < Nase > ?	<input type="checkbox"/>
e	Wie heisst < Kaba >, wenn du statt /a/ immer ein /i/ sagst?	<input type="checkbox"/>

Abbildung 33. Beispiel-Item 6 mit inhaltlicher Nähe zu Berufserfahrungswissen

$\text{Primar}_{\text{mitBE}}$, bei denen die Studienzeit weiter zurückliegt, sind mit dem dahinterstehenden Konzept der phonologischen Bewusstheit möglicherweise weniger vertraut, oder es haben im Verlaufe der Lehrtätigkeit Vergessensprozesse eingesetzt. $\text{Primar}_{\text{mitBE}}$ können aber auf ihr Erfahrungswissen zurückgreifen. Dagegen können $\text{Primar}_{\text{ohneBE}}$ am Ende der Grundausbildung oder SHP am Ende der Zusatzausbildung in Schulischer Heilpädagogik eher aktuelles, formales Lehrbuchwissen aktivieren. Zwar weisen die meisten SHP ebenfalls Berufserfahrung auf, das erworbene, aktualisierte Lehrbuchwissen könnte bei dieser Subgruppe aber stärker im Zentrum stehen als bei Lehrpersonen, die in den vergangenen Jahren vor allem in der Praxis standen. Auch folgendes Beispiel (Abbildung 34) lässt eine solche Interpretation zu.

Kinder mit Leseschwächen kompensieren ihr Defizit manchmal, indem sie ratend lesen. Das Leseverständnis ist dadurch beeinträchtigt. Mit welchen Leseaufgaben finden Sie als Lehrperson heraus, ob das Kind wirklich genau und sinnverstandend liest? Formulieren Sie eine konkrete Leseaufgabe für das Kind, die zum genauen und sinnverstandenden Lesen auf Satzebene zwingt. Das Kind soll dabei nicht laut vorlesen. Falls Sie die Antwort(en) nicht wissen, schreiben Sie „ich weiss es nicht“.
Stille Leseaufgabe für das Kind (umschreiben Sie die Aufgabe nicht, sondern stellen Sie sie direkt)

Abbildung 34: Beispiel-Item 18 mit inhaltlicher Nähe zu Berufserfahrungswissen

Auch hier müssen Lehrpersonen mit Berufserfahrung möglicherweise eher auf ihr Erfahrungswissen zurückgreifen als die übrigen beiden Gruppen. Dieselbe Erklärung liegt für Item 8 (*MC: Identifizieren von Elementen der phonologischen Bewusstheit*) und Item 23 (*Offenes Item: Hilfestellung zur Leseflüssigkeit mit Silbenbogen*) auf der Hand. Die Berufserfahrung führt allerdings nicht dazu, dass die Items besonders gut gelöst werden. Die absolute Lösungshäufigkeit liegt jeweils zwischen derjenigen von Substichprobe *Primar_{ohneBE}* und *SHP*.

Items zu Leseverständnis und Leseflüssigkeit

Die Items 9 und 17 weisen bei den Substichproben *Primar_{mitBE}* und *SHP* ähnliche relative Itemschwierigkeiten auf, sind aber für *Primar_{ohneBE}* im Verhältnis überproportional einfacher als erwartet. Beide Items beziehen sich auf Aspekte des Leseverständnisses und werden von den *Primar_{ohneBE}* anders gelöst als von den übrigen Substichproben. Der Unterschied ist besonders bei Item 17 (Abbildung 35) augenfällig. Allerdings ist eine plausible Erklärung für das differente Lösungsverhalten äußerst schwierig.

Die Vermittlung von Lesestrategien für das Leseverständnis (z.B. Reziprokes Lesen, Herausschreiben von Kernaussagen) ist wichtig. Unter welchen Voraussetzungen ist die Vermittlung von Lesestrategien besonders sinnvoll ? (Eine Antwort ankreuzen)		
Wenn Kinder...		
a	aufgrund lesetechnischer Probleme Mühe haben, sich für die Inhalte zu motivieren.	<input type="checkbox"/>
b	Probleme haben mit dem flüssigen, fehlerfreien Lesen von Texten, die ihrem Alter entsprechen.	<input type="checkbox"/>
c	flüssig lesen, aber Probleme haben, Inhalte zu strukturieren und mit dem Vorwissen zu verbinden.	<input type="checkbox"/>
d	wenig Leseerfahrung haben und Mühe haben, Texte sinnbetont zu lesen.	<input type="checkbox"/>
e	beim Lesen die Wortbedeutungen im mentalen Lexikon nur langsam abrufen können.	<input type="checkbox"/>

Abbildung 35. Beispiel-Item 17 zum Leseverständnis

Auffällig ist, dass bei der Substichprobe *Primar_{ohneBE}* im Vergleich zu den anderen Substichproben auch Distraktor **b** (9.5%) plausibel erscheint. Dies ist insofern bemerkenswert, da auch Item 9 (*Zusammenhang Leseflüssigkeit und Leseschwierigkeit*), bei dem es um Aspekte der Leseflüssigkeit geht, für die Substichprobe *Primar_{ohneBE}* überproportional viel einfacher war als für die anderen beiden Substichproben. Möglicherweise haben beide Items bei den *Primar_{ohneBE}* eine ganz bestimmte Assoziation zum Begriff der Leseflüssigkeit geweckt, die bei den anderen Substichproben nicht vorlag.

Items ohne plausible Erklärung für DIF

Bei Item 5, 19 und 27 entfällt eine plausible Erklärung. So ist etwa bei Item 5 (*MC-Item: Identifizieren einer Leseschwierigkeit auf Wortebene*) unklar, weshalb es im Vergleich der Substichproben $\text{Primar}_{\text{ohneBE}}$ vs. SHP so unterschiedlich funktioniert, während für den Vergleich von Substichprobe $\text{Primar}_{\text{ohneBE}}$ vs. $\text{Primar}_{\text{mitBE}}$ sowie $\text{Primar}_{\text{mitBE}}$ vs. SHP Messinvarianz vorliegt. Dasselbe gilt für Item 27 (*Antwort-Batterie: Phonemanalyse von 6 Begriffen*). Hier muss angenommen werden, dass u.a. Rate-Effekte und weitere unerkannte latente Faktoren die Lösungsmuster beeinflusst haben.

Item 19 (Abbildung 36) ist besonders erwähnenswert, da es in allen *drei* Substichproben deutlich anders funktioniert. Eine Analyse der gewählten Distraktoren erfolgt in Kap. 7.2.2. Die Substichprobe SHP schneidet mit einer Lösungshäufigkeit von 49.1% mit großem Abstand am besten ab während die anderen Substichproben das Item kaum richtig gelöst haben. Für die Substichprobe $\text{Primar}_{\text{mitBE}}$ war das Item noch schwieriger als zu erwarten gewesen wäre.

Um einen Text richtig und ohne Hilfe verstehen zu können, muss ein bestimmter Anteil an Wörtern korrekt gelesen werden. Wie hoch muss dieser Anteil mindestens sein? (Eine Antwort ankreuzen)		
a	50%-60%	<input type="checkbox"/>
b	60%-70%	<input type="checkbox"/>
c	70%-80%	<input type="checkbox"/>
d	80%-90%	<input type="checkbox"/>
e	90%-100%	<input type="checkbox"/>

Abbildung 36. Beispiel-Item 19 mit DIF in allen drei Substichproben

7.2.3.4 Splitten von Items

Als Alternative zum Ausschluss aller Items, die DIF aufweisen, wurde über das Splitten von Items versucht, Messinvarianz zu operationalisieren. Das Vorgehen wird u.a. im Rahmen eines Methodenberichts zum Projekt „*Harmonisierung der obligatorischen Schule*“ (HarmoS) beschrieben (Ramseier, 2008) und in der vorliegenden Studie analog durchgeführt.

In der vorliegenden Studie wurde eine *Basis-Skala* mit zwei alternativen Skalen (*Skala 1* und *Skala 2*) verglichen. Gemeinsame Ausgangslage für die Schätzung der Itemschwierigkeiten in Skala 1 und 2 sind sechs Items, die in den Substichproben keine DIF aufweisen. Für diese sechs Items wurden in allen Substichproben dieselben Itemschwierigkeiten verwendet und für relevant verschieden funktionierende Items je eigene Itemschwierigkeiten geschätzt. Als Kriterien für eine „relevante DIF“ wurde die Differenz der Logit-Beiträge von ≥ 0.43 sowie die signifikante Abweichung der verglichenen Itemschwierigkeiten ($t > 1.96$) festgelegt (Kap. 6.3.3).

Bei *Skala 1* wurden elf Items, die im Gruppenvergleich DIF aufwiesen, in zwei Items gesplittet, Item 19 wurde ganz aufgeteilt. Die elf Items wurden also in einer der drei Substichproben als

separates, in den andern beiden Substichproben als gleiches Item behandelt. Wo genau einer der drei Gruppenunterschiede zu groß war, wurde das Item in jenen beiden Gruppen nicht aufgeteilt, deren Differenzbetrag kleiner war. Ausnahme bildete Item 5, um eine Gleichverteilung der Verbindungen zwischen Gruppen zu erhalten. Damit wurden Items abgeleitet, die nur in Substichproben erhoben wurden, die ursprünglich unter sich keine relevante DIF aufwiesen. Die Items tragen so je zur Schätzung der Personenfähigkeit (WLE) bei, beeinflussen aber den jeweiligen Gruppenunterschied auf der so konstruierten Skala nicht. Dies hat allerdings den Nachteil, dass die Gruppenvergleiche auf unterschiedlichen, zwischen zwei Gruppen gemeinsamen Aufgaben basieren.

Bei *Skala 2* wurden zwölf Items in je drei Items, also vollständig aufgespalten, und die *Basis-Skala* funktionierte ohne Itemsplitting. Der Nachteil von Skala 2 ist, dass der Gruppenvergleich lediglich auf den sechs Items ohne DIF basiert.

In der Folge wurden die Gruppenmittelwerte auf Basis der Personenfähigkeiten-Schätzer (WLE) verglichen. Die Mittelwerte wurden mit verschiedenen Datensätzen (Skala 1 und Skala 2 mit gesplitteten Items und Basis-Skala ohne splitting) produziert (Tabelle 24).

Tabelle 24: Skalenvergleich unter Berücksichtigung von DIF auf Grundlage der WLE-Mittelwerte (Skala₁₈)

Substichproben	Basis-Skala Items nicht gesplittet		Skala 1 ^a Items teilweise gesplittet		Skala 2 ^b Items vollständig gesplittet	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
I Primar _{ohneBE}	-0.45	0.73	-0.51	0.74	-0.46	0.75
II Primar _{mitBE}	-0.08	0.63	-0.23	0.65	-0.29	0.65
III SHP	0.93	0.95	0.61	0.95	0.46	0.93
η^2	0.352		0.258		0.195	

^a11 Items gesplittet und paarweises nicht-splitten von Items in Gruppen mit kleinstem DIF, Item 19 für jede Gruppe gesplittet.

^b12 Items in je drei Items gesplittet, für jede Gruppe einzeln.

Der Vergleich der drei Skalen zeigt, dass sich die WLE-Mittelwerte in allen drei Skalen in derselben Richtung unterscheiden. In allen drei Skalen ist der Mittelwert der Primar_{ohneBE} am niedrigsten, gefolgt von Primar_{mitBE} und schließlich der Substichprobe SHP mit dem höchsten Mittelwert. Die Unterschiede sind zudem über die Skalen hinweg recht ähnlich, so dass mit den verschiedenen Skalen keine grundsätzlich anderen Ergebnisse produziert werden. Die Effektstärken, ausgedrückt durch das Eta-Quadrat (η^2), sind in den Skalen etwas unterschiedlich. Durch Skala 2 können Mittelwertunterschiede der WLE-Werte etwas weniger gut durch die Gruppenzugehörigkeit erklärt werden als durch Skala 1 und die Basis-Skala. Weiter geben die Korrelationen der WLE-Werte der drei Skalen Aufschluss über die „Robustheit“ – mit und ohne Berücksichtigung skalarer Messinvarianz. Die WLE-Werte der Basis-Skala und der Skala 1

korrelieren sehr hoch, $r(443) = .99, p < .001$; dies gilt ebenso für die Basis-Skala und Skala 2, $r(443) = .98, p < .00$.

Insgesamt sprechen diese Werte für eine ausreichende Robustheit der Basis-Skala im Vergleich zu den Skalen 1 und 2, in denen Messinvarianz realisiert wurde. Für die Verwendung der Basis-Skala spricht außerdem, dass Gruppenunterschiede auf einer breiteren inhaltlichen Basis erfasst werden und dass die Inhaltsvalidität besser gewährleistet ist.

7.2.3.5 Fazit aus der differenziellen Itemfunktionsanalyse

Das Befragungsinstrument misst in verschiedenen, nach Berufsbiografie gebildeten Substichproben jeweils unterschiedliche Konstrukte. Das heißt, dass die Messinvarianz für das Befragungsinstrument nicht gegeben ist. Die differenzielle Itemfunktion zeigt sich etwa ähnlich ausgeprägt, jedoch bei unterschiedlichen Items im jeweiligen paarweisen Vergleich der Substichproben. Lediglich sechs Items wurden durchgängig über alle Substichproben hinweg gleich gelöst; ein Item wurde über alle Substichproben hinweg deutlich unterschiedlich gelöst. Für die weitere Verwendung des Instruments bedeutet dies, dass für Gruppenvergleiche in der vorliegenden Art nur Items berücksichtigt werden dürfen, die im Vergleich miteinander keine DIF aufweisen. Eine hohe drop-out-Quote von Items aufgrund von DIF ist nicht ungewöhnlich und auch in anderen Studien (Blömeke et al., 2009; Ramseier et al., 2011; Schindler et al., 2019) zu finden.

Mit dem Ausschluss handelt man sich allerdings massive Einschränkungen hinsichtlich der inhaltlichen Validität des Befragungsinstruments ein (Schulte, Nonte & Schwippert, 2013). Da aufgrund der ausgeprägten DIF ein Ausschluss der betreffenden Items nicht sinnvoll wäre, wurde – wie bei Schulte et al. (2013) vorgeschlagen – ein interpretativer Umgang gewählt.

Da Messinvarianz immer dann relevant ist, wenn Mittelwertvergleiche geplant sind (Schulte et al., 2013), musste für den später in dieser Arbeit durchgeführten Mittelwertvergleich (Validierung durch Gruppenvergleiche, Kap. 7.2.5) ein Umgang mit DIF gefunden werden. Über die Operationalisierung von Messinvarianz durch Itemsplitting konnten verschiedene Skalen produziert und Mittelwerte verglichen werden. Dabei hat sich gezeigt, dass sich die Ergebnisse aus den beiden gesplitteten Skalen nicht grundlegend von denjenigen unterscheiden, die mit der Basis-Skala mit den nicht-gesplitteten 18 Items produziert wurden. Damit konnte eine Entscheidungsgrundlage für die weitere Verwendung der Basis-Skala₁₈ geschaffen werden. Auch hinsichtlich der Inhaltsvalidität ist die Basis-Skala vorteilhafter.

7.2.4 Ergebnisse zur Dimensionalität des Befragungsinstruments

Grundlage für die folgenden Analysen ist das in Kap. 6.3.4 gezeigte Modell. Um zu prüfen, ob die theoretisch angenommenen Dimensionen auch empirisch belastbar sind, werden im Folgenden die Ergebnisse des Devianzdifferenz-Tests (χ^2 -Test) berichtet. *ConQuest*, das für diese Analyse verwendet wurde, erlaubt es, mehrdimensionale Modelle zu schätzen und mit dem eindimensionalen Modell zu vergleichen (Ramseier et al., 2008). Der Test wurde zunächst für Skala₂₈ und anschließend für Skala₁₈ durchgeführt.

7.2.4.1 Dimensionierung des Professionswissens nach Inhalten

In folgenden Tabelle 25 sind die Statistiken des Modells zusammengefasst, bei denen die Items zunächst nach inhaltlichen Dimensionen unterschieden wurden. Gemäß dem theoretischen Modell, auf dessen Grundlage die Items entwickelt wurden, sind 15 Items dem *Grundlagenwissen* und 13 Items dem *diagnostischen Wissen* zugeordnet. In Skala₁₈ liegt eine Verteilung von zehn Items für das *Grundlagenwissen* und acht Items für das *diagnostische Wissen* vor.

Tabelle 25: Modellvergleich Grundlagenwissen und diagnostisches Wissen vs. eindimensionales Modell für Skala₂₈ und Skala₁₈

Skala ₁₈									
Modell	Korrelation ^a	Devianz	Para- meter	<i>n</i>	AIC	BIC	Modellvergleich ^b		
							Differenz der Devianz	Differenz der <i>df</i>	<i>p</i>
Skala ₂₈									
1-dimensional		18 041.005	43	443	18 127	18 155			
2-dimensional	.94	18 038.534	45	443	18 129	18 158	2.47	2	.291
Skala ₁₈									
1-dimensional		10 917.999	27	443	10 972	10 989			
2-dimensional	.95	10 915.445	29	443	10 973	10 992	2.55	2	.279

^a Messfehlerbereinigte Korrelation der latenten Dimensionen. ^b Vergleich mit eindimensionalem Modell; Differenz der Devianz (χ^2 -Test).

Bei Skala₂₈ zeigt sich ein Vorteil des eindimensionalen Modells gegenüber dem zweidimensionalen Modell. Die höheren AIC- und BIC-Werte des zweidimensionalen Modells weisen darauf hin, dass die Abweichung von den Daten größer ist als beim eindimensionalen Modell. Der χ^2 -Test zeigt zudem, dass das zweidimensionale Modell keine bedeutsame Verbesserung gegenüber dem reduzierten Modell bringt.

Auch für Skala₁₈ zeigt sich ein Vorteil des eindimensionalen Modells gegenüber dem zweidimensionalen Modell. Die höheren AIC- und BIC-Werte des zweidimensionalen Modells weisen auch hier darauf hin, dass die Abweichung von den Daten größer ist als beim eindimensionalen Modell. Der χ^2 -Test zeigt auch für Skala₁₈, dass das zweidimensionale Modell keine

bedeutsame Verbesserung gegenüber dem reduzierten Modell bringt. Die Korrelation der beiden latenten Dimensionen ist zudem extrem hoch (sowohl für das Modell von Skala₂₈ wie auch von Skala₁₈).

Insgesamt muss die Annahme einer zweidimensionalen Struktur entlang der Inhalts-Dimensionen verworfen werden. Zumindest bilden die den Dimensionen zugeordneten Items das postulierte Modell nicht ab. Ob eine andere als die theoretisch angenommene Struktur vorliegt und ob Items, anders zugeordnet, auf interpretierbare latente Variablen laden, wird später in Kap. 7.2.4.3 mittels exploratorischer Faktorenanalyse geprüft. Weiter zeigt sich der erwartete inhaltliche Zusammenhang der Dimensionen überaus deutlich. Die sehr hohen Korrelationen der beiden latenten Dimensionen lassen die Interpretation einer eindimensionalen Struktur zu. Da die inhaltliche Nähe der beiden Dimensionen bereits bei der theoretischen Konzeption der Items erkennbar war, überrascht dieses Ergebnis nicht.

7.2.4.2 Dimensionierung des Professionswissens nach kognitiven Anforderungen

Analog zum vorangehenden Kapitel wurden die Items nach kognitiven Anforderungen unterschieden und das Modell entlang dieser Dimensionierung geprüft (Tabelle 26). Hier wurden die Dimensionen *Fachwissen* und *fachdidaktisches Können* unterschieden, die im Ordnungsschema nach Shulman (1986; 1987) dem „content knowledge“ und dem „pedagogical content knowledge“ entsprechen. In Skala₂₈ wurden 17 Items dem *Fachwissen* und elf Items dem *fachdidaktischen Können*. In Skala₁₈ belegen acht Items die Dimension *Fachwissen* und zehn Items die Dimension *fachdidaktischen Können*.

Tabelle 26: Modellvergleich Fachwissen und fachdidaktisches Können vs. eindimensionales Modell für Skala₁₈ und Skala₂₈

							Modellvergleich ^b		
Modell	Korrela- tion ^a	Devianz	Para- meter	<i>n</i>	AIC	BIC	Differenz der Devianz	Differenz der <i>df</i>	<i>p</i>
Skala₂₈									
1-dimensional		18 041.005	43	443	18 127	18 155			
2-dimensional	.98	18 031.248	45	443	18 121	18 150	9.76	2	.008**
Skala₁₈									
1-dimensional		10 917.999	43	443	10 972	10 989			
2-dimensional	.99	10 917.626	45	443	10 976	10 994	0.37	2	.830

^a Messfehlerbereinigte Korrelation der latenten Dimensionen. ^b Vergleich mit eindimensionalem Modell; Differenz der Devianz (χ^2 -Test). **Signifikanz < 0.01

In Tabelle 26 zeigt sich bei Skala₂₈, dass die AIC- und die BIC-Werte für das eindimensionale Modell größer sind als für das zweidimensionale Modell und dass die Werte damit stärker vom

Modell abweichen. Der p -Wert für den χ^2 -Test ist außerdem signifikant, was für eine bessere Modellpassung des zweidimensionalen Modells sprechen würde. Allerdings ist die Korrelation zwischen den latenten Dimensionen so hoch, dass hier dennoch sinnvollerweise lediglich von *einer* Dimension auszugehen ist.

In Skala₁₈ zeigen die AIC- und BIC-Werte sowie der χ^2 -Test, dass das komplexere Modell nicht besser passt als das eindimensionale Modell. Wiederum ist die Korrelation der latenten Dimensionen sehr hoch. Die theoretisch postulierte Mehr-Dimensionalität muss also auch entlang dieser beiden Dimensionen verworfen werden. Auch hier ist die inhaltliche Nähe der latenten Dimensionen naheliegend, so dass dieses Ergebnis ebenfalls wenig erstaunt.

7.2.4.3 Exploratorische Faktorenanalyse

Um nach möglichen weiteren Zusammenhängen der Items und nach Erklärungen latenter Dimensionen zu suchen, wurde unter Verwendung der ML-Schätzmethode mit *M-Plus* eine exploratorische Faktorenanalyse durchgeführt. Dafür wurde die Einstellung für ordinale Daten mit obliquen Rotation vorgenommen. Im Folgenden werden die entsprechenden Ergebnisse dargestellt, wobei nur noch Modelle der Skala₁₈ für die Stichprobe ohne Laien gerechnet werden.

Voraussetzungen der EFA und graphische Inspektion zur Extraktion von Faktoren

Um die Eignung der Daten für eine Faktorenanalyse zu beurteilen, wurden zunächst die Voraussetzungen geprüft. Sowohl der Bartlett-Test ($\chi^2(153) = 985.67.09$ $p < .001$) als auch der KMO-Koeffizient ($KMO = .83$) weisen darauf hin, dass sich die Items insgesamt für die Faktorenanalyse eignen und dass die Voraussetzungen erfüllt sind (Bühner, 2011, 347). Um im ersten Schritt die Anzahl Faktoren zu extrahieren, wurde ein graphischer Scree-Test durchgeführt, und es wurden Faktoren bestimmt, die Eigenwerte > 1 haben (Abbildung 37).

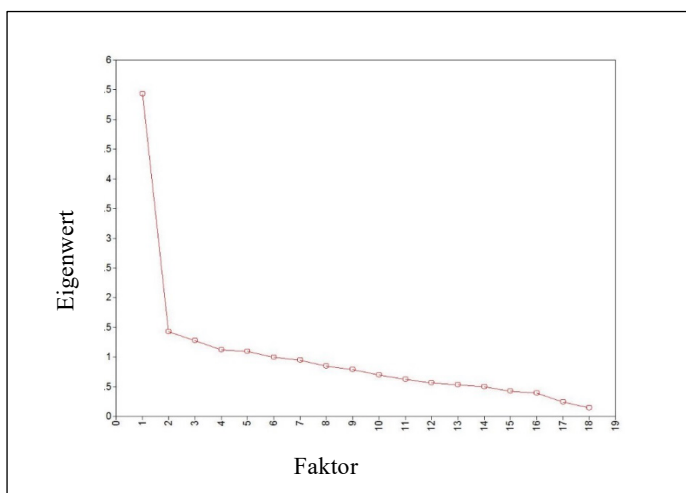


Abbildung 37. Scree-Test zur Extraktion von Faktoren auf Grundlage des Eigenwertkriteriums mit Skala₁₈

Fünf Faktoren weisen einen Eigenwert > 1 auf, wobei hier ein deutlicher Eigenwertabfall nach Faktor 1 zu verzeichnen ist. Tabelle 27 zeigt die Eigenwerte, und erklärt die Varianz der fünf Faktoren im Einzelnen.

Tabelle 27: Eigenwerte, Prozent der erklären Varianz und kumulierte Erklärte Varianz der fünf extrahierten Faktoren, Skala₁₈

Faktor	Eigenwerte	Erklärten Varianz %	Kumulierten Varianz %
1	5.55	30.17	30.17
2	1.38	7.92	38.09
3	1.24	7.09	45.18
4	1.14	6.20	51.38
5	1.06	6.07	57.46

Damit klären die fünf extrahierten Faktoren aufsummiert 57.46% der Gesamtvarianz. Bei Modellen mit weniger Faktoren müssen hier Einschränkungen in der erklärten Varianz in Kauf genommen werden. Als Richtwert schlagen Homburg und Giering (1996 S. 12) einen Mindestanteil von 50% der kumulierten Varianz vor.

Obwohl die Ergebnisse auf das Vorliegen von fünf Faktoren mit Eigenwerten > 1 hinweisen, wurde aufgrund des Scree-Tests (Eigenwertabfall nach Faktor 1) und aufgrund theoretischer Überlegungen die Ein-Faktor- und die Zwei-Faktoren-Lösung weiterverfolgt und inferenzstatistisch ausgewertet. Schon bei einem Modell mit drei Faktoren zeigt sich nämlich, dass der dritte Faktor nur noch durch zwei Items repräsentiert würde und dass eine theoretisch sinnvolle Interpretation einer Drei-Faktoren-Lösung zudem nicht gelinge würde. Faktoren, die durch drei oder weniger Items repräsentiert sind, sind unter dem Gesichtspunkt der Reliabilität selbst bei homogenen Items nicht sinnvoll (Bühner, 2011).

Analyseschritte der exploratorischen Faktorenanalyse

In der folgenden inferenzstatistischen Auswertung werden die Ergebnisse auf einer objektiven Basis mittels χ^2 -Test dargestellt, indem Modelle mit unterschiedlicher Faktorenzahl miteinander verglichen und hinsichtlich ihrer Passung mit Modellen mit mehr Faktoren beurteilt werden (Tabelle 28).

Modelle mit unterschiedlicher Faktorenzahl werden gegeneinander getestet, wobei steigende χ^2 -Werte dahingehend interpretiert werden können, dass Modelle mit mehr Faktoren schlechter passen. Da hier die Modellpassung als Nullhypothese formuliert wird, bedeutet ein signifikantes Ergebnis (Signifikanzniveau 5%), dass dieses abgelehnt wird und ein Modell mit mehr Faktoren besser passt (Bühner, 2011).

Tabelle 28: Vergleich von Faktorenlösungen im Rahmen der exploratorischen Faktorenanalyse

Verglichene Modelle	χ^2	df	p
1 Faktor vs. 2 Faktoren	43.41	17	.000**
2 Faktoren vs. 3 Faktoren	20.91	16	.182
3 Faktoren vs. 4 Faktoren	21.11	15	.134

**< .001

Es zeigt sich somit, dass eine zweidimensionale Lösung einer Einfachstruktur vorzuziehen wäre. Eine Lösung mit mehr als zwei Faktoren scheint aber keinen deutlichen Gewinn mehr zu bringen, da sich der χ^2 -Wert nur unwesentlich ändert.

Um eine inhaltliche Analyse vornehmen zu können, wurden die Faktorladungen für die zweidimensionale Lösung aufgezeichnet, und die Items, welche die inhaltliche Dimension (*diagnostisches Wissen/Grundlagenwissen*) und die Dimension kognitiver Anforderungen (*Fachwissen/fachdidaktischen Können*) repräsentieren, wurden verortet (Tabelle 29). Die Angabe der Itemschwierigkeit auf der Grundlage der KTT diene zudem der Überprüfung eines möglichen „Schwierigkeitsfaktors“, da es denkbar wäre, dass Items gleicher Schwierigkeit in einem Faktor gebündelt wurden (Bühner, 2011).

Tabelle 29: Verortung der Items im zweidimensionalen Modell

Item Nr.	Kurzbeschreibung der Items	Schwierigkeitsindizes	Faktorladung Kommunalität	GW / DW	FW / FDK	
Faktor 1		p_i	λ	h^2		
2	Audio-Item: Analyse von Lesekompetenzen 4. Klasse	.75	.44*	.15	DW	FDK
5	SC-Item: Identifizieren einer Leseschwierigkeit auf Wortebene	.81	.53*	.31	DW	FDK
6	SC-Item: Identifizieren der einfachsten Leseaufgabe auf Wortebene	.72	.46*	.21	GW	FDK
7	Ordnungsaufgabe: Ordnen von Leseaufgaben nach Lesestrategie	.37	.41*	.39	GW	FDW
8	MC-Item: Identifizieren von Elementen der phonologischen Bewusstheit	.36	.33*	.25	GW	FW
12	SC-Item: Typische Symptome von Leseschwierigkeiten im Vergleich zu gut lesenden Kindern	.79	.60	.39	DW	FW
16	Zuordnungsaufgabe: Spezifisches, diagnostisches Potenzial von Leseaufgaben	.87	.35*	.32	DW	FDK
22	Antwortbatterie richtig-falsch: Analyse von Beispielen bei Anlauttabelle	.46	.50*	.18	GW	FW
23	Offene Item: Hilfestellung zur Leseflüssigkeit mit Silbenbogen	.55	.41*	.14	GW	FDK
27	Antwort-Batterie richtig falsch: Phonemanalyse von 6 Begriffen	.29	.47*	.32	GW	FDK
Faktor 2		p_i	λ	h^2		
9	SC-Item: Zusammenhang Leseflüssigkeit und Leseverständnis	.54	.26* ^a	.11	GW	FW
11	SC-Item: Typische Leseschwierigkeiten im Anfangsunterricht	.51	.37*	.18	DW	FW
13	Offene Item: Diagnose mittels 1-Minuten Leseprobe	.54	.47*	.36	DW	FDK
14	Offenes Item: Analyse einer Aufgabe zur Benennungsgeschwindigkeit	.29	.92*	.65	DW	FDK
19	SC-Item: Mindestanteil an korrekt gelesenen Wörtern für unabhängiges Lesen	.16	.81*	.67	GW	FW
28	Offenes Item: Was ist der <i>Lix</i> und über welche Textmerkmale wird er berechnet	.17	.72*	.81	GW	FW
Items, die auf keinen Faktor laden						
17	SC-Item: Vermittlung von Lesestrategien für das Leseverständnis	.74	—	0.13		
18	Offene Item: Entwickeln einer Leseaufgabe für Kinder, die ratend lesen	.14	—	0.16		

Anmerkung. GW = Diagnostisches Wissen, DW = Grundlagenwissen, FW = Fachwissen FDK = Fachdidaktisches Können. Die Faktoren 1 und 2 korrelieren signifikant, $r(443) = .62, p < .05$. ^a Die Faktorladung von Item 9 liegt bei vorliegender Signifikanz knapp unter dem kritischen Wert von $\lambda = .30$.

Die Resultate zeigen, dass auf den ersten Faktor deutlich mehr Items laden als auf den zweiten Faktor. Zudem scheinen Item 17 und Item 18 keinen der beiden Faktoren zu repräsentieren, was sich in geringen Ladungen $< .30$ zeigt. Die Kommunalitäten liegen bei allen Items über dem kritischen Wert von $h^2 > .10$, sind aber bei einigen Items eher niedrig. Damit wird ausgedrückt, dass lediglich etwas über 10% der Varianz der Items 2, 9, 17 und 23 durch den entsprechenden Faktor erklärt wird und dass diese Items nur bedingt durch den Faktor repräsentiert sind. Ein Zusammenhang mit der Itemschwierigkeit kann hier nicht als Begründung dienen.

Auf der Suche nach alternativen Strukturen könnte man von einem Schwierigkeitsfaktor ausgehen; die mittlere Schwierigkeit des ersten Faktors liegt bei $p_i = .60$, jene des zweiten Faktors bei $p_i = .37$. Allerdings laden auch Items mit mittleren Schwierigkeiten auf den zweiten Faktor, so dass dieser Unterschied auch zufällig sein könnte. Auch das Argument eines formatbedingten Faktors ist nicht durchgängig belastbar, obschon drei der fünf offenen Items auf den zweiten Faktor laden. Zwei der offenen Items laden dagegen *nicht* auf diesen Faktor, so dass nicht von einer formatbedingten latenten Struktur auszugehen ist.

Korrelation der latenten Dimensionen

Wie im Folgenden noch gezeigt wird, fällt die inhaltliche Interpretation der explorativ gefundenen Faktoren schwer. Bei aller Vorsicht in Bezug auf die inhaltliche Interpretation ist hier von zwei unterscheidbaren, aber dennoch stark überlappenden Dimensionen auszugehen. Die hohe Korrelation $r = .62$, $p < .05$ der beiden latenten Dimensionen kann als Argument gelten, auch hier von einer Einfachstruktur auszugehen. Die Faktorladungen für das eindimensionale Modell liegen mit Ausnahme von Item 22 ($\lambda = .29$, $p < .05$) bei Skala₁₈ in einem Wertebereich von .30 bis .89 und erreichen damit den kritischen Wert von $\lambda \geq .30$. Eine Übersicht zu den Faktorladungen und Kommunalitäten des eindimensionalen Modells befindet sich im Anhang (Tabelle 38). Beim eindimensionalen Modell würden aufgrund der signifikanten Faktorladungen alle 18 Items im Itempool belassen, da jedes Item einen signifikanten Beitrag zur Erfassung eines gemeinsamen latenten Konstrukts leistet. Allerdings wird durch den einen Faktor lediglich 30.17% der Varianz aufgeklärt und Item 2, 22 und 23 weisen Kommunalitäten von $< .10$ auf. Diese Items werden damit durch diesen Faktor ungenügend erklärt. Damit würden auch bei diesem Modell kritische Items identifiziert und müssten in der vorliegenden Form aus dem Itempool entfernt oder zumindest überarbeitet werden.

7.2.4.4 Inhaltliche Beschreibung der ermittelten Faktoren

Die Items der zweidimensionalen Lösung sind weder entlang der theoretischen Dimensionierung *Grundlagenwissen (GW)* und *diagnostisches Wissen (DW)* noch entlang der Unterscheidung *Fachwissen (FW)* und *fachdidaktisches Können (FDK)* verortet. Aufgrund der Daten muss vielmehr von einer völlig anderen inhaltlichen Struktur ausgegangen werden. Eine inhaltlich sinnvolle Interpretation der alternativen latenten Dimensionen erschließt sich allerdings nicht unmittelbar. Eine vorsichtige Annahme wäre allenfalls beim zweiten Faktor von *ausbildungsspezifischem Wissen* auszugehen. Dieses Wissen wird erst im Rahmen der Zusatzausbildung in Schulischer Heilpädagogik erworben. Auf der anderen Seite repräsentieren die Items des ersten Faktors eher *Basiswissen*, das bereits im Laufe der Grundausbildung vermittelt wird.

Ausbildungsspezifisches Wissen von SHP

Für eher ausbildungsspezifisches Wissen/Können stehen exemplarisch etwa die Items 14 und 28 (Abbildung 38).

Ein Kind in der 2. Klasse soll Farben (Aufgabe 1) und Zahlen (Aufgabe 2) so rasch wie möglich benennen. Welche für das Worterkennen relevante Kompetenz wird mit diesen beiden Aufgaben erfasst?
Falls Sie die Antwort nicht wissen, schreiben Sie „ich weiss es nicht“.

Aufgabe 1

■	■	■	■	■
■	■	■	■	■
■	■	■	■	■
■	■	■	■	■
■	■	■	■	■
■	■	■	■	■

Aufgabe 2

5	6	4	8	2
6	2	8	4	5
2	8	6	5	4
8	4	5	6	2
4	5	8	2	6
5	8	6	2	4

Antwort:

Abbildung 38. Beispiel-Item 14 für ausbildungsspezifische Wissensanforderungen

Bei Item 14 wird nach dem Konzept der Benennungsgeschwindigkeit gefragt, das im Zusammenhang mit der Diagnostik von leseschwachen Kindern nach neueren Erkenntnissen relevant ist (Kap. 2.9.1). Möglicherweise wird dieses Wissen im Rahmen der Grundausbildung zur Lehrperson nicht vermittelt. Das Item wird nämlich von Studierenden ohne Zusatzausbildung in Schulischer Heilpädagogik, d.h. von Primar_{ohneBE} (16.8%) und Primar_{mitBE} (18.8%) seltener richtig gelöst. Der Anteil richtiger Lösungen bei den Studierenden mit Zusatzausbildung in Schulischer Heilpädagogik (SHP) liegt dagegen bei 60.0%. Noch deutlicher zeigt sich diese Differenz bei Item 28, das nach dem Leseschwierigkeitsindex (Lix) fragt (Abbildung 39). Auch hier kann davon ausgegangen werden, dass Analysen zu Oberflächenmerkmalen der Textschwierigkeit (Klicpera et al., 2013, 70ff; Wember, 1999) erst im Rahmen der Zusatzausbildung in Schulischer Heilpädagogik vermittelt werden.

Texte können sich in ihrem "Lix" unterscheiden. Beantworten Sie untenstehende Fragen a) b).
*Bitte **nicht** nur Stichwörter, sondern ganze Sätze schreiben. Falls Sie die Antwort(en) nicht wissen, schreiben Sie „ich weiss es nicht“.*

a	Was bedeutet der Lix?
b	Über welche Textmerkmale wird der Lix berechnet?

Abbildung 39. Beispiel-Item 28 für ausbildungsspezifische Wissensanforderungen

73.9% der Studierenden mit Zusatzausbildung in Schulischer Heilpädagogik erzielen hier einen oder zwei Punkte, während von den anderen Substichproben weniger als 10.0% die teilrichtige oder richtige Antwort geben. Über dieses Wissen verfügen also praktisch nur die SHP.

Auch bei Item 11 (*Typische Leseschwierigkeiten im Anfangsunterricht*), Item 13 (*Diagnose mittels 1-Minuten-Leseprobe*) und Item 19 (*Mindestanteil an korrekt gelesenen Wörtern für unabhängiges Lesen*) ist die Begründung plausibel, dass es sich um ausbildungsspezifisches Wissen handeln könnte. Bei Item 9 (*Zusammenhang Leseflüssigkeit und Leseverständnis*) ist eine solche Interpretation ebenfalls legitim.

Basiswissen

Betrachtet man den Itemkomplex um den ersten Faktor etwas genauer, so muss hier von einem noch heterogeneren latenten Konstrukt ausgegangen werden. Vergleicht man etwa die Items mit den höchsten Ladungen – Item 5 (*Identifizieren einer Leseschwierigkeit auf Wortebene*) und Item 12 (*Typische Symptome von Leseschwierigkeiten im Vergleich zu gut lesenden Kindern*) – sowie Item 22 (*Richtig-falsch-Antwortbatterie: Analyse von Beispielen bei Anlauttafel*) so erschließt sich die Gemeinsamkeit nicht direkt. Die These, dass es sich bei diesen Items um *Basiswissen* handelt, wird allerdings mit großer Vorsicht aufgestellt; dies auch vor dem Hintergrund, dass im Rahmen der exploratorischen Faktorenanalyse – im Gegensatz zur konfirmatorischen Faktorenanalyse – hypothesengenerierend vorgegangen wird (Bortz & Schuster, 2010).

7.2.4.5 Fazit zur Dimensionalisierung des Befragungsinstruments

Die theoretisch postulierten Dimensionen konnten auch in der Hauptuntersuchung nicht bestätigt werden. In der vorliegenden Studie legen die Modellvergleiche den Schluss nahe, von einem zweidimensionalen Konstrukt auszugehen. Allerdings sind die latenten Dimensionen inhaltlich anders organisiert als theoretisch angenommen. Dies zeigt sich deutlich in der exploratorischen Faktorenanalyse, bei der eine vorsichtige Interpretation zum Vorliegen einer Dimension „ausbildungsspezifisches Wissen von SHP“ und „Basiswissen“ vorgenommen wurde. Allerdings soll dies lediglich als Hypothese verstanden werden. Hingegen lässt die hohe und zudem signifikante Korrelation der latenten Dimensionen auch den Schluss zu, von einem eindimensionalen Konstrukt auszugehen. Am ehesten könnte hier von unterschiedlichen, aber dennoch sich überlappenden Wissensbereichen ausgegangen werden. Augenscheinlich handelt es sich um ein sehr heterogenes Konstrukt, was bereits aufgrund der theoretischen Modellierung zu erwarten war.

7.2.5 Validierung: Ausbildungssensitivität und Kontrastgruppenvergleich

Im Folgenden werden die Analyseschritte zur kriterienorientierten Validität berichtet. Die Kriteriumsvalidität wurde anhand der Wissensunterschiede, die durch die Berufsbiografie der drei

Substichproben mitbedingt sind, beurteilt. Zudem wurde ein Wissens-Kontrast zwischen den Substichproben und der Kontrollgruppe (Laien) erwartet.

Für die anschließend durchgeführte einfaktorielle ANOVA und die Kontrastanalyse wurden die erforderlichen Voraussetzungen (Normalverteilung und Varianzhomogenität) geprüft. Bei deren Verletzung wurde über ein entsprechend alternatives Vorgehen entschieden. Datengrundlage waren die Personenfähigkeits-Schätzer (WLE) aus dem Rasch-Modell (Kap. 6.3.1), die aus der finalen Skala¹⁸ hervorgingen. Die Itemschwierigkeiten wurden für die Substichproben I, II, III gemeinsam geschätzt. Diese Skala wurde anschließend zur nachträglichen Schätzung der Fähigkeiten der Laien verwendet, so dass sich das Wissen der Laien auf der Grundlage der Skala der Substichproben I, II, III interpretieren lässt.

Da 36 Laien das Item 2 nicht vorgelegt bekamen, wurden diese als „Missing“ behandelt. Bei der Schätzung der Personenfähigkeit floss diese Information mit ein, so dass die Werte für alle 64 Laien geschätzt wurden. Für die folgenden Analysen lag entsprechend eine Stichprobe von $N = 507$ Studienteilnehmenden vor. Die Analysen wurden mit den WLE-Werten aus dem Rasch-Modell durchgeführt (Begründung Kap. 6.3.1).

7.2.5.1 Haupteffekte des Vergleichs der drei Substichproben und der Kontraststichprobe

Mit Ausnahme der Kontraststichprobe der Laien weichen die Werte der drei Substichproben von einer Normalverteilung ab (*Kolmogorov-Smirnov Signifikanzkorrektur nach Lilliefors*). Auch der Levene-Test zeigt signifikante Unterschiede in den Varianzen ($p = .001$) zwischen den vier Gruppen. Für die ANOVA wurden deshalb das robuste Verfahren Brown-Forsythe-Test und der Welch-Test angewendet.

In Abbildung 40 wird der Wissensanstieg der drei Substichproben und der Kontraststichprobe (Laien) in einem Boxplot dargestellt. Der querschnittliche Anstieg des Wissens/Könnens zeigt sich hier in erwarteter Richtung von den Laien ($M = -0.87$, $SD = 0.64$, $n = 64$) über die $\text{Primar}_{\text{ohneBE}}$ ($M = -0.45$, $SD = 0.73$, $n = 190$) zu den $\text{Primar}_{\text{mit BE}}$ ($M = -0.08$, $SD = 0.63$, $n = 139$) und schließlich zur kompetentesten Gruppe der SHP ($M = 0.93$, $SD = 0.95$, $n = 114$).

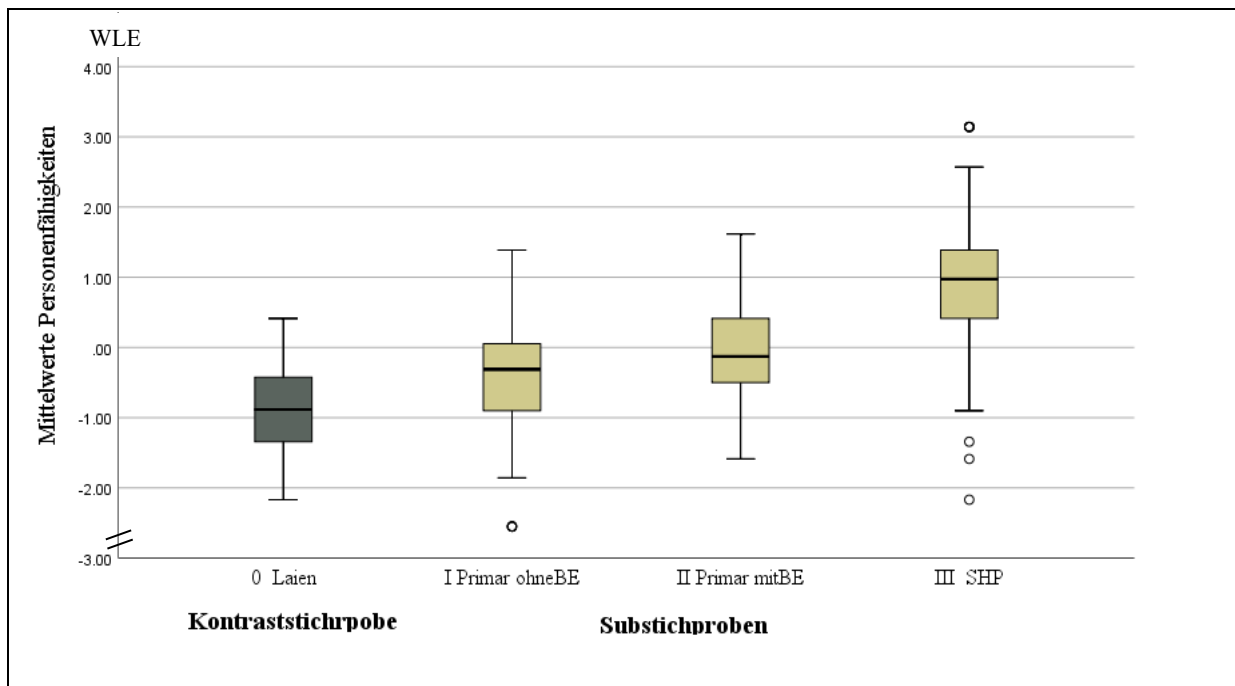


Abbildung 40. Vergleich des querschnittlichen Wissensanstiegs auf Grundlage der Personenfähigkeiten-Schätzer der Substichproben und Kontraststichprobe ($N = 507$).

Die Graphik zeigt, dass insbesondere in der Substichprobe der SHP vier Personen wider Erwarten gut/schlecht abgeschnitten haben, in Substichprobe Primar_{ohneBE} betrifft dies eine Person. Diese Ausreißer dürften das Ergebnis aber nicht wesentlich beeinflusst haben, weshalb diese Probandinnen und Probanden in der Stichprobe belassen wurden.

Im Vergleich zu den übrigen Substichproben zeigt sich der deutlichste Wissensanstieg bei der Substichprobe der SHP. Die ANOVA mit robusten Verfahren ergibt einen signifikanten Haupteffekt für den Faktor Gruppenzugehörigkeit, und zwar unabhängig davon, ob die Laien mit einbezogen wurden – Welch-Test $F(3, 221.47) = 92.751, p < .01$, Brown-Forsythe $F(3, 379.15) = 115.30, p < .01, \eta^2 = 0.399$ – oder die Laien nicht berücksichtigt wurden – Welch-Test $F(2, 253.31) = 90.12, p < .01$, Brown-Forsythe $F(2, 311.08) = 112.45, p < .01, \eta^2 = 0.352$. Eta-Quadrat liegt bei Ausschluss der Substichprobe der Laien in einem hohen Bereich, unter Einbezug der Laien noch etwas höher. Die Effekte können damit als substantiell eingestuft werden (Döring & Bortz, 2016, S. 821). Die Gruppenzugehörigkeit erklärt hier 35.2% der Wissensunterschiede zwischen den Substichproben. Dieses Ergebnis soll im Folgenden weiter differenziert werden, um genauere Aussagen zur Bedeutsamkeit dieser Unterschiede machen zu können.

7.2.5.2 Einzelvergleich der drei Substichproben und Kontraststichprobe

Für die Einzelvergleiche wurde jede Gruppe mit allen weiteren Gruppen verglichen. Im Folgenden sollen aber nur die interessierenden Vergleiche (Vergleich mit jeweils benachbarter Gruppe) erwähnt werden (Tabelle 30). Um die ungleichen Varianzen und Stichprobengrößen

zu berücksichtigen, wurde für den Einzelvergleich der Games-Howell-Test herangezogen (Field, 2013). Um die Effektstärken der Gruppenunterschiede zu ermitteln, wurde Cohens d berechnet. Dafür wurden die Mittelwertdifferenzen der Personenfähigkeits-Schätzer (WLE) und die gepoolten Standardabweichungen der zu vergleichenden Gruppen verwendet (Pissarek & Schilcher, 2017). Die unterschiedlichen Gruppengrößen wurden in der Berechnung der gepoolten Standardabweichung berücksichtigt und die Effektstärke wurden um einen positiven Bias korrigiert (Döring & Bortz, 2016).

Tabelle 30: Gruppenunterschiede zwischen Substichproben und Laien-Kontraststichprobe, Games-Howell-Test

Einzelvergleich der Substichproben		Mittlere Differenz der WLE	p	95%-Konfidenzintervall unten oben		Effektstärke d
0 Laien	I Primar _{ohneBE}	0.413	.000	0.18	0.64	0.60
I Primar _{ohneBE}	II Primar _{mitBE}	0.375	.000	0.18	0.57	0.55
II Primar _{mitBE}	III SHP	1.011	.000	0.74	1.28	1.29

Der post-hoc-Einzelvergleich zeigt, dass der Wissensunterschied für alle Vergleiche auf dem 1%-Niveau signifikant wird. Auch mit dem konservativeren Test mit Bonferroni-Korrektur (Eid et al., 2015) erreichen alle Unterschiede eine Signifikanz auf dem 1%-Niveau. Die kleinste Mittelwert-Differenz zeigt sich zwischen den Substichproben Primar_{ohneBE} und Primar_{mitBE}. Allerdings ist auch hier ein mittlerer Effekt zu verzeichnen (Döring & Bortz, 2016). Die Berufserfahrung scheint sich etwas weniger im gemessenen Wissen/Können niederzuschlagen als etwa das Wissen, das im Rahmen der Zusatzausbildung in Schulischer Heilpädagogik erworben wird. Auffallend hoch fällt der Effekt zwischen den Substichprobe Primar_{mitBE} und SHP aus. Studierende am Ende der Ausbildung in Schulischer Heilpädagogik verfügen mit Abstand über ein höheres Wissen/Können als Primarlehrpersonen mit Berufserfahrung. Der Wissensunterschied zwischen der Laien-Kontraststichprobe und der Stichprobe Primar_{ohneBE} drückt sich in einer mittleren Effektstärke aus.

7.2.5.3 Fazit zur kriterienorientierten Validität

Grundsätzlich kann festgehalten werden, dass das Instrument sensitiv auf ausbildungsbiografische Unterschiede reagiert. Auch eine Abgrenzung zwischen Professionswissen und Wissensbeständen der Laien kann bestätigt werden, auch wenn hier ein deutlicheres Ergebnis wünschenswert gewesen wäre. Die Berufserfahrung scheint sich etwas weniger im gemessenen Professionswissen niederzuschlagen als das Wissen, das im Rahmen der Zusatzausbildung in Schulischer Heilpädagogik erworben wird. Sehr viel stärker als die Berufserfahrung wirkt sich aber der ausbildungsspezifische Hintergrund auf das gemessene Wissen aus. Die Validitätsanalyse zeigt, dass Studierende am Ende der Ausbildung in Schulischer Heilpädagogik ein mit Abstand

höheres Professionswissen besitzen als Primarlehrpersonen mit Berufserfahrung. Insgesamt vermag das Befragungsinstrument diese Gruppenunterschiede sensitiv zu erfassen und erfüllt damit ein zentrales Validitätskriterium.

7.3 Die Ergebnisse der Hauptstudie im Überblick

Im Folgenden werden die Ergebnisse der verschiedenen Analysen aus Kap. 7.2 zusammengefasst, um eine Gesamteinschätzung der Qualität der verwendeten Items sowie des gesamten Befragungsinstruments zu ermöglichen. Aus dem ursprünglichen Befragungsinstrument der Hauptstudie mit 28 Items wurden zehn Items mit unbefriedigenden statistischen Kennwerten ausgeschlossen (Kap. 7.2.1). Nachfolgend werden nur die 18 der insgesamt 28 Items einbezogen, die in der Hauptstudie statistisch befriedigende Kennwerte ergaben. Die Analysen basieren auf den Daten einer Stichprobe von Studierenden der Primarstufe ohne Berufserfahrung (Primar_{ohneBE}, $n = 190$), Studierenden mit Berufserfahrung (Primar_{mitBE}, $n = 139$) und Studierenden der schulischen Heilpädagogik (SHP, $n = 114$).

7.3.1 Zusammenfassung der Distraktorenanalyse, Item-Skalenanalyse und Itemfunktionsanalyse

Um die Qualität des Befragungsinstrumentes zu beurteilen, wurden verschiedene Analysen auf Item- und Skalenebene durchgeführt. Dabei wurden zentrale Gütekriterien der Klassischen Testtheorie (Trennschärfe, Itemschwierigkeit, interne Konsistenz [Cronbachs α]) sowie Gütekriterien auf der Grundlage des zur Item-Response-Theorie gehörenden Rasch-Modells (Weighted Meansquare [MNSQ], differenzielle Itemfunktion [DIF]) überprüft. Für diese theoriebasierten Gütekriterien wurden Grenzwerte definiert, die in Kap. 6.3 begründet und in Tabelle 31 angegeben sind. Als weiteres Qualitätskriterium wurde berücksichtigt, ob die Distraktoren der Items bei gebundenem Format von den Studienteilnehmenden ausreichend oft gewählt wurden, um als plausibel eingestuft werden zu können.

Die Übersicht in Tabelle 31 zeigt, ob bzw. wie die einzelnen Items die Kriterien der Klassischen Testtheorie (KTT) und der Item-Response-Theorie (IRT) erfüllen. In der Tabelle werden die wichtigsten Itemkennwerte zusammengetragen. Für jedes Item wird festgehalten, ob die fünf Kriterien: 1) *plausible Distraktoren*, 2) *Trennschärfe*, 3) *Schwierigkeitsindizes*, 4) *MNSQ/T* und 5) *DIF* erfüllt sind. In der letzten Spalte wird zusammenfassend angegeben, wie viele Gütekriterien das einzelne Item erfüllt.

Tabelle 31: Übersicht der Werte auf Itemebene der Skala₁₈ nach KTT und IRT (Rasch-Modell)

Item Nr.	Kriterium Distraktorenanalyse	Kriterien KTT		Kriterien IRT (Rasch-Modell)			erfüllte Kriterien
	Plausible Distraktoren	Trennschärfe r_{it}	Schwierigkeits- indizes p_i	MNSQ / T	DIF Paarweiser Vergleich Substichprobe I-II-III		0-5
2	a b c d e	.23	.75	1.09 1.3	keine DIF		3
5	c e	.30	.81	0.96 -0.6	moderat		4
6	alle plausibel	.28	.72	1.01 0.2	moderat/hoch		3
7	alle plausibel	.47	.37	0.96 -0.8	moderat		5
8	alle plausibel	.38	.36	1.01 0.3	moderat		5
9	alle plausibel	.22	.54	1.06 1.9	moderat/hoch		3
11	a b c e	.27	.51	1.03 0.8	keine DIF		3
12	a b d e	.36	.79	0.93 -1.1	keine DIF		4
13	offenes Item	.42	.54	1.02 0.4	keine DIF		5
14	offenes Item	.40	.29	0.92 -1.7	hoch		4
16	alle plausibel	.33	.87 max.	0.98 -0.1	keine DIF		4
17	alle plausibel	.22	.74	1.03 0.5	hoch		3
18	offenes Item	.23	.14 min.	1.00 0.0	hoch		2
19	alle plausibel	.43	.16	0.88 -1.5	hoch		4
22	alle plausibel	.24	.46	1.12 2.2	moderat		3
23	offenes Item	.22	.55	1.05 1.5	hoch		3
27	anderes Format	.38	.29	1.03 0.2	moderat		5
28	offenes Item	.55	.17	0.81 -2.3	hoch		4
Mittelwerte		.33	.50				

Anmerkung. Distraktoren: Als plausibel wurden Distraktoren eingestuft, wenn sie zumindest von Laien > 5% gewählt wurden, selbst wenn sie von den Substichproben I, II, III teilweise < 5% gewählt wurden. DIF: Falls im Vergleich der Substichproben I vs. II, I vs. III und II vs. III keine oder höchstens eine moderate DIF bestand, wurde das Kriterium als „erfüllt“ gewertet. Die Bewertung der übrigen Kriterien richtet sich nach den in Kap. 6.3 erwähnten Grenzwerten: $r_{it} \geq .30$ / $p_i = .15 - .85$ / $MNSQ > 0.80$ und < 1.20 / $T -2$ bis $+2$.

Eine Einschätzung der Qualität der einzelnen Items ist davon abhängig, wie viele Kriterien einbezogen, wie die Kriterien gewichtet und welche Grenzwerte festgelegt werden (Moosbrugger & Kelava, 2020). Bei den hier vorliegenden Analysen wurden eher liberale Grenzwerte verwendet. Weitgehend befriedigend sind die Werte zur Itemschwierigkeit und die MNSQ-Werte. Problematisch – und die Qualität des Befragungsinstrumentes einschränkend – sind insbesondere Items mit niedrigen Trennschärfen sowie die hohe DIF.

Abschließend lässt sich feststellen, dass trotz eher liberaler Grenzwerte (Anmerkung Tabelle 31) nur vier Items *alle* Kriterien erfüllen, weitere sechs Items erfüllen die Kriterien zumindest teilweise.

Neben der Analyse auf Itemebene sind Werte der Gesamtskala zu beachten. Es wurden wiederum Werte auf Grundlage beider Testtheorien (KTT, IRT) berechnet. In Tabelle 32 werden die Skalenwerte der Gesamtstichprobe und anschließend der drei Substichproben zusammengefasst.

Tabelle 32: Übersicht der Werte auf Skalenebene der Skala₁₈ nach KTT und IRT

	<i>N</i>	KTT			IRT (Rasch-Modell)		
		Punkte (max. 26) M (SD)	Trennschärfe r_{it} (Durchschnitt)	Cronbachs α	Varianz	EAP-Reli- abilität	WLE-Re- liabilität
Gesamtstichprobe	443	12.84 (4.61)	.33	.75	0.65	.78	.76
Substichproben							
I Primar _{ohneBE}	190	10.47 (3.57)	.21	.60	0.33	.62	.61
II Primar _{mitBE}	139	12.36 (3.30)	.17	.52	0.22	.53	.52
III SHP	114	17.31 (4.30)	.34	.75	0.76	.77	.74

Anmerkung. Die Einschätzung auf Skalenebene richtet sich nach den in Kap. 6.3 erwähnten Grenzwerten: $r_{it} \geq .30$ / Cronbachs $\alpha > .70$ / Varianz > 0.60 / EAP- und WLE-Reliabilität $> .60$.

Die Skalenwerte für die Gesamtstichprobe sind insgesamt befriedigend, auch wenn sie die kritischen Werte teilweise nur knapp erreichen. Anders sieht die Situation aus, wenn die Werte getrennt nach Substichprobe betrachtet werden. In der Substichprobe Primar_{ohneBE} liegen außer der EAP- und WLE-Reliabilität alle Werte unter dem festgelegten Grenzwert; in der Substichprobe Primar_{mitBE} müssen ausnahmslos alle Werte als unbefriedigend eingestuft werden. In der Substichprobe SHP sind die Ergebnisse hingegen alle zufriedenstellend. Damit zeigt sich abschließend, dass das Instrument je nach Stichprobe, Daten unterschiedlicher Qualität liefert und ein stichprobenbezogener Pretest für die künftige Anwendung zu empfehlen ist.

7.3.2 Zusammenfassung der Ergebnisse zur Dimensionierung des Befragungsinstruments

Zur Beurteilung der Validität des Instrument wurde geklärt, welche Struktur bzw. Dimensionalität das latente Konstrukt „Professionswissen“ aufweist. Zum einen wurde konfirmatorisch geprüft, ob die theoretisch postulierte Struktur, die der Entwicklung des Befragungsinstruments zu Grunde lag, empirisch belegt werden kann. Zum anderen wurde exploratorisch untersucht, ob das Professionswissen möglicherweise in anderen als den angenommenen Strukturen verankert ist.

Um die Frage zu klären, ob das Professionswissen wie im theoretischen *Modell des lesebezogenen Wissens/Könnens* (Kap. 4.1) angenommen, organisiert ist oder eine eindimensionale Struktur des Wissens vorliegt, wurde jeweils ein zweidimensionales Modell mit einem eindimensionalen Modell verglichen. Der Vergleich zwischen eindimensionalem und zweidimensionalem Modell wurde einmal für die *Inhaltsdimension* (Grundlagenwissen / Diagnostisches Wissen) und einmal für die Dimension *Kognitive Anforderung* (Fachwissen / Fachdidaktisches Wissen/Können) durchgeführt. Mittels Devianzdifferenz-Test wurde der Modell Fit der

komplexeren zweidimensionalen Struktur gegenüber der reduzierten eindimensionalen Struktur überprüft und statistisch mit dem χ^2 -Test abgesichert.

Die Ergebnisse des χ^2 -Devianzdifferenz-Test (Kap. 7.2.4) zeigen Folgendes: Modellvergleich *Grundlagenwissen* und *diagnostisches Wissen* versus eindimensionales Modell für Skala₁₈, $p(443) = .279$, Modellvergleich *Fachwissen* und *fachdidaktisches Können* versus eindimensionales Modell für Skala₁₈, $p(443) = .830$. Die nicht signifikanten Differenzwerte ($p > .01$) sowie die niedrigeren AIC- und BIC-Werte der eindimensionalen Modelle zeigen in beiden Vergleichen an, dass die zweidimensionalen Modelle keine bedeutsame Verbesserung gegenüber den eindimensionalen Modellen bringt (Gäde et al., 2020). Das Fazit aus der Analyse zur Unterscheidbarkeit der theoretisch postulierten Dimensionen ist, dass sich weder die Subdimension *Grundlagenwissen* vom *diagnostischen Wissen* noch die Subdimension *Fachwissen* vom *fachdidaktischen Wissen* empirisch trennen ließ.

Die exploratorische Faktorenanalyse wurde durchgeführt, um allfällige alternative Strukturen zur angenommenen Unterscheidung von *Grundlagenwissen* und *diagnostischem Wissen* bzw. *Fachwissen* vom *fachdidaktischen Wissen* zu finden. Die folgende Tabelle 33 zeigt das Ergebnis der exploratorischen Faktorenanalyse für die zweidimensionale sowie die eindimensionale Lösung.

Tabelle 33: Übersicht der exploratorischen Faktorenanalyse, eindimensionales und zweidimensionales Modell

Zweidimensionales Modell					Eindimensionales Modell	
Faktor 1 Basiswissen			Faktor 2 Ausbildungsspezifisches Wissen			
Item Nr.	Faktorladung λ	Kommunalität h^2	Faktorladung λ	Kommunalität h^2	Faktorladungen λ	Kommunalität h^2
2	.44*	.15			.30*	.09 kritisch
5	.53*	.31			.51*	.27
6	.46*	.21			.40*	.16
7	.41*	.39			.61*	.38
8	.33*	.25			.50*	.25
9			.26* kritisch	.11	.32*	.10
11			.37*	.18	.41*	.17
12	.60	.39			.57*	.33
13			.47*	.36	.59*	.35
14			.92*	.65	.65*	.42
16	.35*	.32			.56*	.32
17	--	--	--	--	.35*	.13
18	--	--	--	--	.39*	.15
19			.81*	.67	.76*	.58
22	.50*	.18			.29*	.08 kritisch
23	.41*	.14			.32*	.10
27	.47*	.32			.54*	.29
28			.72*	.81	.89*	.80

Korrelation der Faktoren, $r(443) = .62$, $p < .05$

Anmerkung. Item 17 und 18 wiesen für die zweidimensionale Lösung zu geringe Faktorladungen und Kommunalitäten auf. Als Grenzwert für die Beurteilung des Beitrags der Faktorladungen wurde der in Kap. 6.3.4 begründete Wert von $\lambda \geq .30$ und $h^2 > .10$ festgelegt.

Zusammenfassend zeigt sich, dass sowohl das eindimensionale als auch das zweidimensionale Modell grundsätzlich empirisch belastbar und inhaltlich interpretierbar ist (Kap. 7.2.4.3 / 7.2.4.4). Eine Analyse der Items im zweidimensionalen Modell lässt die vorsichtige Interpretation zu, dass es sich beim ersten Faktor um Inhalte handelt, die man als *Basis-Wissen* bezeichnen könnte und beim zweiten Faktor um Inhalte, die zum *ausbildungsspezifischen Wissen* gehören. Ausbildungsspezifisches Wissen wird im Rahmen der Zusatzausbildung in Schulischer Heilpädagogik erlernt, das Basis-Wissen wird dagegen bereits in der Grundausbildung zur Primarlehrperson erworben. Allerdings ist diese Unterscheidung inhaltlich vorsichtig zu interpretieren und als Hypothese zu verstehen.

Für beide Modelle gilt einschränkend, dass jeweils zwei Items aus dem Itempool entfernt werden müssen, da die Kommunalitäten oder Faktorladungen in den Modellen die kritischen Werte unterschreiten (Bühner, 2011, S. 358). Die Kommunalitäten liegen im eindimensionalen Modell bei Item 2 und 22 unter dem festgelegten Wert, im zweidimensionalen Modell sind die Faktorladungen bei Item 17 und 18 zu gering und repräsentieren keinen der Faktoren. Alle vier Items weisen zusätzlich niedrige Trennschärfen auf (Tabelle 31) und müssten für die weitere Verwendung überarbeitet oder aus dem Itempool entfernt werden.

7.3.3 Zusammenfassung der Ergebnisse des Kontrastgruppenvergleichs und zur Ausbildungssensitivität

Zur kriterienorientierten Validierung wurde untersucht, ob sich in den drei Substichproben (Primar_{ohneBE}, Primar_{mitBE} und SHP) Wissensunterschiede zeigen, die durch die unterschiedlichen Berufsbiographien zu erklären sind (Ausbildungssensitivität). Die Analyse, ob das Befragungsinstrument in der Lage ist das Professionswissen unterschiedlicher Substichproben zu unterscheiden, wurde unter Einbezug der Laienstichprobe ($n = 64$) durchgeführt. Die Laienstichprobe diente als Kontraststichprobe, von der erwartet wurde, dass sie in der Befragung deutlich schlechter abschneidet als die übrigen Substichproben.

Das Ergebnis der ANOVA zeigt unter Einbezug der Laien einen signifikanten Haupteffekt von $\eta^2 = 0.399$ für den Faktor Gruppenzugehörigkeit (Kap. 7.2.5.1). Die Gruppenzugehörigkeit erklärt damit 39.9% der Wissensunterschiede zwischen den Substichproben. Der Effekt ist als starker Effekt zu bezeichnen (Döring & Bortz, 2016, S. 821).

Der Einzelvergleich der vier Substichproben zeigte einen Wissensanstieg in erwarteter Richtung: Laien weisen das geringste Wissen auf, gefolgt von den Primar_{ohneBE} über die Primar_{mitBE} bis hin zu den SHP, die über das grösste Professionswissen verfügen. Im Einzelvergleich der Substichproben inklusive Laien wurden folgende Effektstärken ermittelt: Laien vs. Primar_{ohneBE} $d = 0.60$, Primar_{ohneBE} vs. Primar_{mitBE} $d = 0.55$, Primar_{mitBE} vs. SHP $d = 1.29$. Die Effektstärken

liegen damit in einem mittleren – und im Falle des Vergleichs mit der Substichprobe der SHP – in einem hohen Bereich. Damit kann dem Befragungsinstrument attestiert werden, dass es ausbildungsbezogene Unterschiede im Professionswissen messen kann. Zudem lässt sich ableiten, dass sich das Wissen von Laien deutlich vom Wissen der drei Substichproben mit Ausbildung und Berufserfahrung unterscheidet. Damit wird bestätigt, dass das Befragungsinstrument seine Funktion insgesamt erfüllt, also Professionswissen misst.

8 Diskussion

Im Folgenden werden zunächst die zentralen Befunde zu den in Kap. 5.1 formulierten Forschungsanliegen diskutiert. Danach folgen Limitationen des Befragungsinstruments, die sich in theoretischer wie auch methodischer Hinsicht ergeben. Abschließend werden Überlegungen zum Ertrag des vorliegenden Projekts für die Forschung und Praxis dargelegt.

8.1 Diskussion des Befragungsinstruments

Ausgehend von theoretischen Herleitungen zur Operationalisierung des Professionswissens und der Übertragung auf den Bereich Lesen (Kap. 4) wurde in der vorliegenden Dissertation ein Befragungsinstrument entwickelt, welches das Wissen und Können zu Grundlagen des Leserwerbs und zur Diagnostik bei Leseschwierigkeiten erfasst. Wie der aktuelle Stand der Forschung zeigt, wurde damit eine Forschungslücke im Bereich der Kognitionsforschung geschlossen, (Kap. 1). Ein zentrales Anliegen war die Anschlussfähigkeit des entwickelten Befragungsinstruments an aktuelle Instrumente, die sich an die Typologie zum Professionswissen von Lehrpersonen von Shulman anlehnen, die im Rahmen verschiedener Studien (Kap. 3.2 und Kap. 3.3) differenziert wurde. Der Bereich *Lesen* wurde bisher nur wenig bearbeitet; angesichts der Bedeutsamkeit des Lesen-Könnens sind Instrumente für diese Domäne aber dringend nötig.

Der entscheidende Beitrag des Instruments liegt darin, dass auf *veränderbare proximale* Faktoren fokussiert wurde und damit auf eine Ebene, die über die Gestaltung der Lehrerinnen- und Lehrerbildung *mitgesteuert* werden kann. Um fundierte Aussagen zur Qualität des erworbenen Wissens und Könnens machen zu können, müssen valide und reliable Befragungsinstrumente vorliegen, die zudem in Durchführung und Auswertung ökonomisch sind. Entsprechende Ziele (Kap. 5.1) wurden für das entwickelte Befragungsinstrument formuliert und die Zielerreichung wurde in einem mehrschrittigen Prozess im Rahme einer Vorstudie und einer Hauptstudie umgesetzt. Inwiefern diese Ziele hinsichtlich ökonomischer Durchführung, objektiver Auswertung, Datenstruktur und Skalierung sowie inhaltlicher Validität und Ausbildungssensitivität erreicht wurden, wird nachfolgend entlang der in Kap. 5.1 formulierten Forschungsanliegen und -ziele zusammengefasst und diskutiert. Anschließend wird auf den Nutzen und die Grenzen des Befragungsinstruments für Forschung und Praxis verwiesen.

8.1.1 Objektivität des Befragungsinstruments und ökonomische Durchführung

Ein zentrales Anliegen war die Entwicklung eines Befragungsinstruments, das in seiner Durchführung und Auswertung objektiv ist und zudem als Gruppenverfahren ökonomisch eingesetzt werden kann. Damit ist folgendes Forschungsziel angesprochen: *Das Befragungsinstrument soll in seiner Durchführung und Auswertung objektiv sein.*

Objektivität und Auswertungsnorm

In der Tradition der Kognitionsforschung wird Kompetenz bzw. Professionswissen als erlernbare kognitive, kontextspezifische Leistungsdisposition verstanden (Zlatkin-Troitschanskaia & Kuhn, 2010, 26). Ausgehend von einem solchen Verständnis (Kap. 3.1) liegt die Entwicklung eines hochstrukturierten testähnlichen Befragungsinstruments zum Professionswissen nahe, wie es im vorliegenden Projekt realisiert wurde. Konkret schlägt sich der Objektivitäts-Anspruch zum einen im formalen Format nieder, zum anderen aber auch darin, dass Wissen und Können *direkt* erfasst werden und nicht über Selbsteinschätzungen bzw. Kompetenzerwartungen (z.B. Studie Schmidt & Schabmann, 2016) oder die Erfassung subjektiver Theorien (z.B. Sodogé, 2010). Im europäischen Raum wurde dieser objektive Ansatz zur Kompetenzmessung bei Lehrpersonen erstmals im Rahmen der COACTIV-Studie umgesetzt und war Anstoß für zahlreiche weitere objektive Messinstrumente (Kunter & Klusmann, 2010). Erfasst wurde dabei, was angehende und ausgebildete Lehrpersonen tatsächlich wissen und nicht, was sie *meinen* zu wissen. Bei der Erfassung von Fachkompetenzen bzw. Fachwissen über subjektive Selbst- oder Fremdeinschätzungen ist mit verschiedenen verzerrenden Effekten (u.a. soziale Erwünschtheit) zu rechnen (Frey, 2006), was bei alternativen Verfahren wie dem vorliegenden Befragungsinstrument nicht zu erwarten ist. Objektivität bezieht sich hier entsprechend auf die direkte quantifizierbare Erfassung kognitiver Wissens- und Könnensbestände, die frei von subjektiven Selbsteinschätzungen erhoben wurden.

Durch die Verwendung von Items mit vorwiegend gebundenem Antwortformat schlägt sich die Durchführungs- und Auswertungsobjektivität auf einer ganz anderen Ebene auch formal nieder (Kap. 5.3.2). Bei gebundenen Formaten treten in der Regel hinsichtlich Auswertungsobjektivität keine Schwierigkeiten auf (Rost, 2004). Im vorliegenden Modell waren nur gerade fünf Items als offene Fragen konzipiert, was vor allem für die Auswertung nach besonderen Maßnahmen verlangt. Hier wurde eine exakte Kodieranweisung erstellt und die Übereinstimmung von drei Raterinnen bzw. Ratern berechnet. Insgesamt kann für die offenen Items von einer zufriedenstellenden Auswertungsobjektivität ausgegangen werden, da die Übereinstimmungsreliabilität nach mehreren Überarbeitungen der Anweisungen im Bereich bei $\kappa = .72$ bis $.89$ lag (Kap. 5.3.4, Tabelle 7). Um eine objektiviertere Durchführung zu gewährleisten, wurden ein Instruktionsmanual erstellt und genaue Angaben zum Ausfüllen des Fragebogens gemacht.

Zu diskutieren ist auch die *Interpretationsobjektivität*, die als weiteres Gütekriterium berücksichtigt werden muss (Moosbrugger & Kelava, 2012a). Dieses Kriterium wurde bei den Zielsetzungen nicht explizit genannt. Überlegungen dazu sind jedoch wichtig für die Diskussion des Befragungsinstruments, da der Aspekt der Auswertungsnorm berührt wird. Die meist durch

eine statistische Norm festgelegte Interpretationshilfe wird in der Regel aus den Ergebnissen einer Eichstichprobe gewonnen und ermöglicht den Vergleich von Testpersonen mit relevanten Bezugsgruppen (Goldhammer & Hartig, 2012). In der vorliegenden Untersuchung wurden keine Prozentrangnormen berechnet, die einen statistischen Vergleichsmaßstab zur Testwertinterpretation ermöglicht hätte (ebd.). Auch wurden keine Schwellenwerte festgelegt, die eine objektivierte Aussage darüber zulassen würde, ab wann eine Person über „genügendes Basiswissen“ oder „Zusatzwissen“ verfügt. Aus den Daten lässt sich lediglich ablesen, dass Primarlehrpersonen am Ende der Ausbildung durchschnittlich 40% der Punkte erreichten, während der Punkte-Anteil bei Primarlehrpersonen mit Berufserfahrung bei 48% und bei SHP bei 67% lag (Kap. 7.2.1.1, Tabelle 16). Ob dies einer festgelegten Erwartung entspricht, könnte sich neben einer statistischen Norm auch über eine kriterienorientierte Interpretation feststellen lassen (Döring & Bortz, 2016). Hier müsste aufgrund theoretischer, inhaltlicher Überlegungen und Normen von Expertinnen und Experten und/oder gemessen an einem Außenkriterium – wie z.B. die Leseleistung der Schülerinnen und Schüler – definiert werden, ab welcher Punktzahl die Mindestanforderungen erfüllt wären (ebd.). Die in der vorgelegten Studie beigezogene Expertinnen- und Expertennorm stützten lediglich den inhaltlichen Rahmen, indem die Aufgaben als grundsätzlich relevant und lösbar bestätigt wurden. Es wurde aber darauf verzichtet, ein quantitatives Maß zu bestimmen, das von angehenden und ausgebildeten Lehrpersonen im Sinne einer *Mindestanforderung* zu erfüllen wäre. Diese Überlegungen knüpfen an die Diskussion um „Standards der Lehrerinnen- und Lehrerbildung“ (Oser, 1997; Oser, Curcio & Duggeli, 2007) an und betreffen die Frage, welche Wissensbestände notwendigerweise erworben werden müssen, um einem handlungsorientierten Gütemaßstab standzuhalten. Im Gegensatz zur Kompetenz – im Sinne einer Fähigkeit in einem Bereich – fungieren Standards darüber hinaus als Maßstab für den erreichten Ausprägungsgrad von Kompetenz und lassen eine Beurteilung (z.B. „unzureichend“) zu (Terhart, 2005). In diesem Sinne sind Standards „nichts anderes als erwartete Kompetenzen“ (Helmke, 2012, S. 142). Es geht damit um eine normative Festlegung davon, was jemand zwingend mitbringen muss, um einen Beruf – hier den Lehrberuf – ausüben zu können (Helmke, 2017). Damit ist eine an Standards orientierte Kompetenzmessung mit einer Form der Normierung verbunden, die mit einem Perspektivenwechsel von der Forschung hin zur Bildungspolitik einhergeht (Helmke, 2012). Dies führt zur Frage, *welches* Professionswissen von Absolventinnen und Absolventen der Lehrerinnen- und Lehrerausbildung erwartet werden kann, damit sie den beruflichen Anforderungen gewachsen sind, da davon ausgegangen wird, dass das Professionswissen von Lehrpersonen erlernt werden kann (Terhart, 2005). Diese

Frage lässt sich hier nicht abschließend beantworten, sie muss aber bei der Interpretation der Ergebnisse mit einbezogen werden.

Zur objektivierbaren Einstufung wäre auch der Vergleich mit einem relevanten Außenkriterium denkbar. Im deutschsprachigen Raum liegen jedoch bisher im Bereich Lesen kaum Untersuchungen vor, die den Zusammenhang des Professionswissens mit anderen Unterrichtsvariablen wie z.B. Leistungsfortschritte untersuchen, so dass Aussagen zur Bedeutsamkeit nur vermutet werden können (Lessing-Sattari & Wieser, 2018). Eine objektiv begründete Interpretationsnorm, etwa durch die Festlegung von Wissensniveaus oder Basisstandards (Ramseier et al., 2011), steht für das entwickelte Befragungsinstrument noch aus und wäre im Zuge der Weiterentwicklung z.B. durch geeignete Außenkriterien oder die empirische Berechnung von Niveaustufen zu begründen.

Ökonomisches Gruppenverfahren

Ein weiteres Anliegen war die Entwicklung eines ökonomischen Verfahrens, das dem späteren Verwendungszweck dienen würde. Sollen später z.B. Vergleichsstudien zur Wirksamkeit der Lehrerinnen- und Lehrerausbildung im hier fokussierten Bereich Lesen durchgeführt werden, wären ökonomisch einsetzbare Befragungsinstrumente unabdingbar. Entsprechend wurde folgendes Ziel formuliert: *Das Befragungsinstrument soll in seiner Durchführung ökonomisch sein und sich als Gruppenverfahren eignen für Lehrpersonen mit unterschiedlichem Vorwissen. bedingt durch Ausbildung und/oder Berufserfahrung.*

Dieser Anforderung wurde durch die Testlänge und das Antwortformat Rechnung getragen. Als kritische Grenze der Testlänge bei Lehrpersonen werden 45 bis 60 Minuten angegeben (Wiprächtiger-Geppert et al., 2015, S. 297). Um die Probandinnen und Probanden nicht über Gebühr zu belasten, wurde die Testlänge im vorliegenden Fall auf 60 Minuten begrenzt, was zahlreiche Konsequenzen hat. Inhaltlich bedeutet dies, den Wissensbereich der *Förderung* von Kindern mit Leseschwierigkeiten weitgehend außer Acht zu lassen. Der Aspekt der „Zumutbarkeit“ (Moosbrugger & Kelava, 2012a, S. 22) setzte also auch inhaltliche Grenzen.

In formaler Hinsicht hatte der Anspruch an die ökonomische Durchführbarkeit zur Folge, dass vorwiegend geschlossene Antwortformate verwendet werden, was auch hinsichtlich der Auswertungsökonomie vorteilhafter ist (Rost, 2004). Im Spannungsfeld verschiedener Ansprüche müssen bei der Entwicklung von Befragungsinstrumenten Entscheidungen getroffen werden, die hier zu diskutieren sind. Argumente der Ökonomie dürfen dabei nicht die einzige Grundlage für die Ausgestaltung eines Instrumentes sein und müssen immer im Verhältnis zu den übrigen Anforderungen – vor allem zum diagnostischen Erkenntnisgewinn – eines Instruments stehen

(Moosbrugger & Kelava, 2012a). Zu klären ist, inwiefern die in dieser Studie vorwiegend geschlossenen Antwortformate, die sich zudem nur begrenzt nach konkreten Unterrichtssituationen ausrichten, zu Lasten der Validität ausfallen (Rost, 2004). Diese Frage stellt sich vor dem Hintergrund, dass offene kontextnahe Fall- und Unterrichtsvignetten häufig als besonders geeignet beschrieben werden, um fachdidaktische, prozeduralisierte Wissensanteile zu erfassen (Brühwiler et al., 2017; König, 2015; Rutsch et al., 2018c; Vogelsang & Reinhold, 2013). Kontextnähe ist allerdings nicht nur mit offenen Formaten zu erreichen. Sie kann durchaus auch über geschlossene und damit meist ökonomischere Formate erreicht werden (Rutsch et al., 2018c). Entscheidend ist also nicht (nur) das Format, sondern auch der Anspruch der *Kontextualisierung*. Insbesondere werden Ansprüche der Validität durch die kontextnahe Erfassung scheinbar eher eingelöst, sofern die Disposition (Wissen, Können) mit der Anforderungssituation (Unterrichtssituation) verbunden ist (König, 2015).

Der Anspruch einer kontextualisierten Erfassung fachdidaktischen Könnens wurde im vorliegenden Projekt durch die vier verwendeten Audio-Items, die als Fallvignetten konzipiert sind, explizit berücksichtigt. In der eigenen Untersuchung zeigte sich allerdings, dass die vier Audio-Items trotz geschlossenem Format etwa ein Drittel der zur Verfügung stehenden Zeit in Anspruch nahmen. Dieser deutliche Nachteil hinsichtlich Durchführungsökonomie wird durch den Vorteil eines praktischen Anforderungsbezugs (Validität) nur über starke inhaltliche Gründe und/oder gute Testkennwerte aufgewertet. Inhaltlich gesehen wären weitere Anstrengungen zur Erfassung des diagnostischen Könnens anhand von Audiobeispielen wünschenswert, da sich diese Fähigkeit letztlich an realen Beispielen widerspiegeln muss. Im vorliegenden Fall waren aber neben der ökonomisch schlechten Bilanz auch die statistischen Testkennwerte (Reliabilität) der Audio-Items mit einer Ausnahme unbefriedigend ($r_{it} < .20$), wodurch die Eignung dieser Items zusätzlich in Frage gestellt werden muss. Diese Beobachtung steht in Einklang mit der Aussage von Blömeke, König, Suhl, Hoth und Döhrmann (2015b, S. 312), dass zu bearbeitende unterrichtsnahe Vignetten meist komplexer sind als andere Antwortformate, was zu einer Reduktion an Aufgaben und damit zu Verlusten an diagnostischer Genauigkeit, Generalisierbarkeit und Reliabilität führt.

Im Zusammenhang mit Fragen der ökonomischen Durchführung wird der Blick abschließend auf die Items mit offenem Antwortformat gerichtet. Die Verwendung der fünf offenen Items kann trotz des höheren Bearbeitungs- und Auswertungsaufwands gerechtfertigt werden, da die ermittelten Trennschärfen befriedigend ($r_{it} > .20$) bis gut ($r_{it} > .40$) sind und die Validität durch offene Reaktionsmöglichkeiten meist als höher einzuschätzen ist als bei vorgegebenen Antworten (Rost, 2004). Insgesamt kann das entwickelte Befragungsinstrument in seiner aktuellen

Fassung mit 18 Items sowohl hinsichtlich Durchführungs- als auch Auswertungsaufwand als ökonomisch eingestuft werden, allerdings zum Preis einer eher handlungsferneren Erfassung des Wissens und Könnens. Hinsichtlich des späteren Verwendungszwecks wird dieser Nachteil in Kauf genommen, da nicht zu erwarten ist, dass aufwändig durchzuführende Instrumente später auch eingesetzt werden – etwa im Rahmen von Forschungsprojekten oder zum Vergleich von Ausbildungsinstitutionen.

8.1.2 Psychometrische Kriterien des Befragungsinstruments

Die Untersuchung psychometrischer Eigenschaften des Befragungsinstruments war ein weiteres zentrales Anliegen des vorliegenden Projektes. Das diesbezügliche Ziel lautet: *Das entwickelte Befragungsinstrument soll unter Anwendung der KTT und der IRT skaliert und auf seine psychometrischen Eigenschaften hin überprüft werden. Namentlich sind dies Itemschwierigkeit, Reliabilität und Trennschärfe.*

Mit dem Einbezug der beiden Messtheorien *Klassische Testtheorie* (KTT) und *Item-Response Theorie* (IRT) wurden Vorteile der jeweiligen Messtheorien (Kap. 6.1.2) genutzt, und es wurde die Empfehlung aufgegriffen, die beiden psychometrischen Theorien zu kombinieren (Blömeke et al., 2015a, 10). Im vorliegenden Projekt wurden neben klassischen Item- und Skalenanalysen auch Skalierungen auf Grundlage des Rasch-Modells durchgeführt.

Einschätzung des Instruments aufgrund zentraler statistischer Kennwerte: Skala₂₈

Zunächst soll ein Blick auf einige Werte aus der Rasch-Skalierung geworfen werden. Grundsätzlich lässt sich feststellen, dass die vorliegenden Daten der Hauptuntersuchung auf der Grundlage der MNSQ-Werte (> 0.80 und < 1.20) mit kleineren Einschränkungen (Kap. 7.2.1.2) eine befriedigende Passung zum Rasch-Modell belegen. Allerdings liegt die mittels Rasch-Modell ermittelte Varianz unter Einbezug aller 28 Items (Skala₂₈) mit einem Wert von 0.35 in einem unbefriedigenden Bereich (Tabelle 13). Dies bedeutet, dass die Items das Fähigkeitsspektrum nicht befriedigend abbilden. Die EAP- und WLE-Reliabilität liegt mit einem Wert von 0.74 in einem befriedigenden Bereich (Kap. 7.2.1.1, Tabelle 19).

Die globalen Einschätzungen der statistischen Kennwerte auf Grundlage der KTT (Kap. 7.2.1.1) zeigen eine akzeptable Abdeckung der Itemschwierigkeiten. Hingegen zeichnet sich hinsichtlich der Trennschärfe ein kritisches Bild ab, was zur Entscheidung führte, trotz eines liberalen cut-off Werts ($r_{it} < .20$), zehn Items für die weiteren Analysen auszuschließen. Als wichtiges Beurteilungskriterium wird in der KTT oft die Trennschärfe aufgeführt (Bühner, 2011). Für die Skalierung mittels Rasch-Modell spielt dieses Kriterium hingegen keine bedeutende Rolle, da hier von identischen Trennschärfe für jedes Item ausgegangen wird (Geiser &

Eid, 2010). Auch wenn gelegentlich davor gewarnt wird, eine feste Untergrenze für die Trennschärfe zu setzen und inhaltliche Überlegungen stärker zu gewichten (Bühner, 2011), wurde diese Selektion unter empirischem Blickwinkel als sinnvoll erachtet, da die Repräsentativität dieser Items für die Gesamtskala zu gering war. Bei einigen der ausgeschlossenen Items waren auch die ermittelten T-Werte auf Grundlage der Rasch-Skalierung kritisch (Kap. 7.2.1.2, Tabelle 17), die Items waren somit in doppelter Hinsicht ungeeignet.

Das Problem geringer Trennschärfen und hoher drop-out-Quoten hat sich wiederholt auch im Rahmen anderer Untersuchungen gezeigt (Ramseier et al., 2011; Riegler & Wiprächtiger-Gepert, 2016). Diskutiert werden muss, inwiefern der Ausschluss von Items die inhaltliche Validität einschränkt. Es stellt sich die Frage, ob das reduzierte Itemrepertoire der Skala₁₈ das zugrunde gelegte Modell dennoch repräsentiert bzw. an welchen Stellen durch die Neukonstruktion von Items inhaltliche Lücken zu füllen sind (reduziertes Modell im Anhang, Tabelle 39). Daraus ergibt sich, dass vor allem im Bereich „diagnostisches Wissen“ zu Facette 1 (*Wissen über Symptome von Leseschwierigkeiten*) und Facette 2 (*Wissen über Leseverläufe bei unauffällig und schwach lesenden Kindern / Wissen über typische Schwierigkeiten bei Stufenübergängen*) sowie im Bereich Grundlagenwissen zu Facette 1 (*Wissen über linguistische Merkmale der deutschen Schriftsprache und Aspekte der Textschwierigkeit*) und Facette 2 (*Wissen über Prozessmodelle des Lesens und Determinanten der Lesekompetenz*) Überarbeitungen oder Neukonstruktionen von Items notwendig wären.

Einschätzung des reduzierten Befragungsinstruments: Skala₁₈

Skalen- und Itemwerte der Skala₁₈ zeigen sich in vieler Hinsicht als brauchbarer als die Werte der Skala₂₈. Mit kleinen Einschränkungen kann dem Instrument (Skala₁₈) bescheinigt werden, wesentliche psychometrische Kriterien zu erfüllen – dies gilt zumindest unter Berücksichtigung der ungeteilten Stichprobe, $N = 443$ (Kap. 7.3.1). Namentlich deckt das Instrument das Fähigkeitsspektrum angehender und ausgebildeter Lehrpersonen bzw. SHP angemessen ab und ist in der Lage, auch im oberen und unteren Leistungsbereich zu differenzieren (Kap. 7.2.1.3, Abbildung 24). Da die Itemschwierigkeits- und Personenfähigkeitsparameter unter Geltung des Rasch-Modells stichprobenunabhängig sind, ist die Itemschwierigkeit einzelner Items über verschiedene Stichproben hinweg generalisierbar (Hambleton, Swaminathan & Rogers, 1991). Dies ermöglicht es, die Schwierigkeiten in den Substichproben direkt zu vergleichen (Ramseier et al., 2011). Dieser Vergleich ergibt, dass die Substichprobe der SHP mit Berufserfahrung die Items generell mit einer höheren Wahrscheinlichkeit zu lösen vermögen als Primarlehrpersonen mit und ohne Berufserfahrung ($\text{Primar}_{\text{ohneBE}}$, $\text{Primar}_{\text{mitBE}}$). Weiter zeigt sich, dass das

Fähigkeitsspektrum der SHP besser abgebildet wird als in den übrigen zwei Substichproben (Kap. 7.2.1.3, Abbildung 24). Die Varianz als quantitatives Maß dafür, wie gut das Wissensspektrum durch die Items abgedeckt wird, kann für die gesamte Stichprobe der Skala₁₈ mit einem Wert von 0.65 als ausreichend akzeptiert werden (Kap. 7.3.1, Tabelle 32). Die Varianz liegt damit in einem vergleichbaren Bereich wie z.B. im Deutschtest der TEDS-LT-Studie (Stancel-Piatak et al., 2013, 34f.). Deutliche Einschränkungen zeigen sich bei getrennter Betrachtung der Substichproben. Insbesondere vermag das Befragungsinstrument das Fähigkeitsspektrum der Primar_{ohneBE} und Primar_{mitBE} nicht hinreichend abzudecken, hingegen weist das Instrument für die SHP eine zufriedenstellende Varianz von 0.76 auf.

Für die ungeteilte Stichprobe erweist sich die Skala₁₈ als reliabel im Sinne interner Konsistenz (Cronbachs $\alpha = .75$) und ausgedrückt durch die EAP-Reliabilität (.78) und WLE-Reliabilität (.76) (Kap. 7.3). Die Werte sind nicht beeindruckend, liegen aber in einem vergleichbaren Wertebereich wie bei anderen raschskalierten Messinstrumenten zum professionellen Wissen von Studierenden (Buchholtz & Kaiser, 2013; Seifert & Schaper, 2012). In der Skala zum lesedidaktischen Wissen von Lehrpersonen in der Studie von Rutsch (2016) liegt die EAP-Reliabilität z.B. bei .59. Eher niedrigere interne Konsistenzen (Cronbachs α) sind auch im von Wiprächtiger-Geppert et al. (2015) entwickelten Instruments im Bereich orthographischen Wissens zu verzeichnen. Die Autorinnen sehen den Grund im heterogenen Aufgabentyp und darin, dass unter dem Terminus „fachdidaktisches Wissen“ oft ein heterogenes Bündel an Wissen und Können subsumiert wird (ebd., S. 287f.). Beide Argumente dürften auch für das vorliegende Befragungsinstrument gelten. Angesichts der Neukonstruktion des Befragungsinstruments sowie der Heterogenität des Konstrukts kann die interne Konsistenz des Instruments insgesamt dennoch als zufriedenstellend eingestuft werden.

Eignung des Befragungsinstruments für die einzelnen Substichproben: Skala₁₈

Gesamthaft zeigen die getrennten Item- und Skalenanalysen nach Substichproben, dass das Instrument (Skala₁₈) sowohl auf der Grundlage der KTT als auch des Rasch-Modells vor allem für die Stichprobe der SHP befriedigende Kennwerte liefert. Dies betrifft die Itemschwierigkeit, (siehe Verteilung in Abbildung 24), die Reliabilität ausgedrückt durch Cronbachs $\alpha = .75$ / EAP-Reliabilität = .77 / WLE- Reliabilität = .74, die Trennschärfen $r_{it \text{ Durchschnitt}} = .34$ und die Varianz = 0.76 (Kap. 7.3.1, Tabelle 32). Die Eignung des Befragungsinstruments ist insbesondere für die Substichprobe Primar_{mitBE} aufgrund problematischer Kennwerte in Frage gestellt. Sowohl Cronbachs $\alpha = .52$ als auch die EAP-Reliabilität = .53 / WLE-Reliabilität = .52 und die Varianz = 0.22 sind unbefriedigend. Auch unter Ausschluss der wenig trennscharfen Items –

also für Skala₁₈ – bleibt das Problem bestehen, es zeigt sich aber in Skala₂₈ noch deutlicher (Kap. 7.2.1, Tabelle 16 und Tabelle 19). Die Frage nach den möglichen Ursachen für die unterschiedlichen Skalenreliabilitäten und die Abdeckungen des Wissensspektrums stellt sich hier erneut, allerdings bezogen auf Unterschiede, die in den Substichproben und nicht in der Heterogenität des Konstrukts (siehe oben) zu suchen sind (Roters, Nold, Haudeck, Kessler & Stancel-Piatak, 2011). Eine Hypothese ist, dass sich die Stichproben in ihren Lerngelegenheiten und Vorerfahrungen deutlich unterscheiden. Möglich ist auch, dass bei der Substichprobe Primar_{mitBE} teilweise eher zufälliges Wissen vorhanden ist und dass die Studierenden dieser Gruppe eher geraten haben (Bremerich-Vos et al., 2011). Diese Erklärung scheint plausibel, da sich diese Substichprobe aus Lehrpersonen mit unterschiedlich langer Berufserfahrung zusammensetzte und dass dadurch das im Studium erworbene Wissen sehr unterschiedlich präsent war. Außerdem wurde das Wissen an unterschiedlichen Ausbildungsinstitutionen erworben. Dagegen war die Gruppe der SHP hinsichtlich des aktuell erworbenen Wissens am Ende der Ausbildung homogener. Dazwischen lag die Substichprobe Primar_{ohneBE}, die bezüglich aktuell erworbener Ausbildungsinhalte relativ homogen sein dürfte. Aufgrund des geringeren Wissens dürften aber auch Rateeffekte eine Rolle gespielt haben.

8.1.3 Struktur des Professionswissens

Ausgehend von Shulmans Typologisierung des Professionswissens und daraus abgeleiteten Kompetenzmodellen wurde in der vorliegenden Studie ein Modell entwickelt, bei dem die beiden inhaltlichen Dimensionen *Grundlagen zum Leseerwerb* und *Diagnostik bei Leseschwierigkeiten* sowie die beiden kognitiven Anforderungen *Fachwissen* und *fachdidaktisches Wissen/Können* unterschieden wurden. Das Forschungsziel bezüglich der Struktur des Befragungsinstruments lautet: *Das entwickelte Befragungsinstrument soll hinsichtlich der postulierten Dimensionen des theoretisch abgeleiteten Modelles zum Professionswissen überprüft werden*

Fehlender Nachweis der Trennbarkeit der postulierten Dimensionen

In Bezug auf die Dimensionen *Fachwissen* und *fachdidaktisches Können* wurde in der vorliegenden Studie der Versuch unternommen die beiden Fertigkeiten zu schärfen und mit kognitiven Anforderungen, wie sie in der TEDS-LT-Studie (Blömeke et al., 2011) operationalisiert sind, zu koppeln (Kap. 4.3.1 und Kap. 4.3.2). Damit wurde dem Aspekt Rechnung getragen, dass mit zunehmendem fachdidaktischem Können immer auch eine Vertiefung des Verständnisses von Fachwissen einhergeht (Neuweg, 2014). Als Erstes wurde die Struktur mittels eines Modellvergleichs mit einem Devianz-Test überprüft, und es wurde jeweils ein zweidimensionales Modell mit einem eindimensionalen Modell verglichen. Die Ergebnisse (Kap. 7.2.4)

zeigen, dass die Annahme zweier unterscheidbarer Dimensionen zum *Grundlagenwissen* und *diagnostischen Wissen* verworfen werden muss. Auch die postulierten Dimensionen *Fachwissen* und *fachdidaktisches Können* ließen sich in Skala₁₈ nicht trennen. Die Dimensionen weisen zudem hohe korrelative Zusammenhänge auf, $r = .95$ (Grundlagenwissen und diagnostisches Wissen) und $r = .99$ (Fachwissen und fachdidaktisches Können), so dass auch auf Grundlage dieses Ergebnisses nicht von unterscheidbaren Dimensionen auszugehen ist. Daraus lässt sich schließen, dass die Dimensionen eng miteinander verzahnt sind und dass eher von einer eindimensionalen Skala zu einem heterogenen Konstrukt auszugehen ist. Die Ergebnisse schließen an die widersprüchliche Forschungslage anderer Studien an. Während die Daten der COACTIV-Studie (Krauss et al., 2011) und der FALKO-Studie (Krauss et al., 2017a) die Hypothese unterscheidbarer, aber sich dennoch überlappender Wissensbereiche stützen, liegen in den Forschungsarbeiten der Michigan-Gruppe (Hill et al., 2004) Hinweise vor, dass es sich bei Fachwissen und fachdidaktischem Wissen um einen gemeinsamen Wissenskomplex handelt. In der TEDS-LT-Studie wurden die Wissensbereiche Fachwissen und fachdidaktisches Wissen getrennt skaliert und entsprechend in getrennten Tests erfasst (Stancel-Piatak et al., 2013). Im empirisch geprüften Modell zum Professionswissen von Lindmeier (2011) wurden die beiden Wissensbereiche bereits in der theoretischen Konzeption aufgelöst und als Komponente des Basiswissens vereint. Auch die Daten der Studie von Rutsch (2016, S. 145) weisen darauf hin, dass sich der Wissenskomplex im Bereich Lesen als eindimensionales Konstrukt beschreiben lässt und dass ein hoher Zusammenhang zwischen den Dimensionen *Förderwissen/Intervention* und *Diagnose/Inhaltswissen* besteht (Korrelation, $r = .87$).

Alternative Strukturen des Professionswissens

Aufgrund des fehlenden Nachweises der theoretisch postulierten Dimensionen wurde nach alternativen Strukturen des Professionswissens gesucht und damit ein weiteres Forschungsziel verfolgt: *Beim entwickelten Befragungsinstrument soll geprüft werden, ob andere Dimensionen als die theoretisch postulierten vorliegen oder ob das Professionswissen als eindimensionales Konstrukt strukturiert ist.*

Zunächst zeigt sich, dass sowohl in der Vor- als auch in der Hauptuntersuchung eine ungleiche Verteilung der Anzahl Items, die auf die beiden Faktoren mit dem höchsten Eigenwert laden, vorliegt. Die latenten Dimensionen werden also durch die Items unterschiedlich gut repräsentiert. Die mittels exploratorischer Faktorenanalyse ermittelte erklärte Varianz liegt bei 31.88% (Voruntersuchung) und 30.17% (Hauptuntersuchung) und steigt unter Berücksichtigung eines zweiten Faktors auf 42.94% (Voruntersuchung) bzw. 38.09% (Hauptuntersuchung) an. Dieser

Anteil muss als eher gering eingestuft werden (Homburg & Giering, 1996, S. 12). Der in der Hauptuntersuchung durchgeführte χ^2 -Modellvergleich zeigt eine plausible Zweifaktorenlösung, die sich bei aller Vorsicht auch inhaltlich interpretieren lässt. Zum einen lässt sich ein Faktor identifizieren, der als *Basiswissen* beschrieben werden könnte, zum anderen ein Faktor, der tendenziell *ausbildungsspezifisches Wissen* abbildet, das erst im Rahmen der Zusatzausbildung zur SHP erworben wird (Kap. 7.2.4.4). Da die Korrelation der Faktoren $r = .62$, $p < .05$ hoch ist, ließe sich aber auch hier die Annahme eines *eindimensionalen* heterogenen Konstrukts rechtfertigen. In jedem Fall berührt das hier operationalisierte Konstrukt verschiedene Bereiche des Leseunterrichts, was auch die teilweise niedrigen Faktorladungen erklärt (Cole, Perkins & Zelkowitz, 2016), die im eindimensionalen Modell teilweise um den kritischen Wert von $\lambda = .30$ (Kap. 7.3.2, Tabelle 33) liegen. Auch die bereits erwähnte Heterogenität der Formate dürfte eine Rolle spielen.

Die vorgefundene Datenstruktur – insbesondere die nicht bestätigte Dimensionierung von Fachwissen und fachdidaktischem Wissen – wirft die Frage auf, weshalb die theoretisch angenommene Struktur durch die Items nicht abgebildet wurde. Eine plausible Erklärung liegt in den Unschärfen der theoretischen Zuordnung von Wissenskomponenten (Facetten) zu Wissensbereichen (Lessing-Sattari & Wieser, 2018). Die Abgrenzung von Fachwissen und fachdidaktischem Wissen erweist sich als nicht immer einfach, „und die Modellierung dieser Wissensbereiche unterscheidet sich in den einzelnen Studien durchaus“ (ebd., S. 45). Wie in Kap. 3.3 dargelegt wurde, waren für die formale Festlegung der relevanten Wissensbestände Rückbezüge auf andere Domänen (u.a. Orthographie) notwendig. Die Schwierigkeit zeigte sich darin, diese Strukturen auf die Domäne Lesen zu übertragen und Inhalte verschiedenen Dimensionen plausibel zuzuordnen. Die Unschärfen dürften nicht nur für die Abgrenzung von Fachwissen und fachdidaktischem Wissen eine Rolle gespielt haben, sondern auch für die Unterscheidbarkeit des Grundlagenwissens vom diagnostischen Wissen. Offensichtlich greifen diese Inhalte stark ineinander über und können nicht in getrennten Subskalen erfasst werden.

Shulmans Postulat von zwei zu differenzierenden, aber sich überschneidenden Wissensbereichen (Fachwissen und fachdidaktisches Wissen) wurde in der FALKO-Studie für alle untersuchten Disziplinen empirisch belegt und ist folglich domänenunabhängig feststellbar (Lindl & Krauss, 2017). Allerdings zeigt sich, dass die empirische Abgrenzbarkeit der Wissensbereiche auch von den untersuchten Fächern abhängt und nicht für alle Fächer gleichermaßen zutrifft (Kap. 3.3.2). Für den Bereich Lesen weist einzig die Studie von Phelps und Schilling (2004) auf die Trennbarkeit von Fachwissen und fachdidaktischem Wissen hin. Es stellt sich die Frage, warum dies im vorliegenden Befragungsinstrument nicht gelungen ist. Hier ist zu diskutieren,

inwiefern die fehlende Unterscheidbarkeit – wie erwähnt – in der Operationalisierung der Konstrukte selbst liegt. Diese Erklärung scheint plausibel, da sich Fachwissen und fachdidaktisches Wissen im Sinne höherer Korrelationen annähern, je anforderungsnäher und je weniger ausbildungsorientiert sie operationalisiert sind (Neuweg, 2014). Dies kann als Beleg dafür gewertet werden, dass die Konstrukte bzw. ihr Verhältnis stark von der jeweiligen Operationalisierung abhängen. Trotz des Versuchs, die Dimensionen Fachwissen und fachdidaktisches Wissen/Können durch die Koppelung mit kognitiven Anforderungen zu schärfen, ist es nicht gelungen, das Wissen vom Können abzugrenzen bzw. entsprechende Skalen zu bilden. Als Folge davon lässt sich mit dem vorliegenden Befragungsinstrument auch nicht prüfen, ob das theoretische Wissen für den Aufbau des Könnens bei Lehrpersonen bedeutsam ist und inwiefern ersteres für das zweite tatsächlich Bedingung ist (Neuweg, 2014).

Weiter können Gründe für die Unschärfe der Dimensionen Fachwissen und fachdidaktisches Wissen/Können möglicherweise auch in den Ausbildungsstrukturen liegen, und zwar in dem Sinne, dass fachdidaktisches Wissen und Fachwissen in der Lehrerinnen- und Lehrerbildung auf Primarstufe nicht getrennt vermittelt werden (Oser, 2015). Interessant sind in diesem Zusammenhang Ergebnisse der COACTIV-Studie, die zeigen, dass die empirische Trennbarkeit von Fachwissen und fachdidaktischem Wissen nicht unabhängig von der Schulform beantwortet werden kann, sondern mit den ausbildungsbezogenen Lehrgelegenheiten zusammenhängt (Krauss et al., 2008). So lag beispielsweise bei Gymnasiallehrpersonen (die ein universitäres Fachstudium absolvieren) die latente Korrelation von Fachwissen und fachdidaktischem Wissen nahe 1, während sich die Kategorien bei nichtgymnasialen Lehrpersonen deutlicher unterscheiden (Krauss et al., 2011). Der positive Zusammenhang von Fachwissen und fachdidaktischem Wissen wird hier als Beleg für den Expertiseansatz gesehen, der eine stärkere Vernetzung von Wissensarten bei spezialisierten Expertinnen und Experten postuliert (ebd.). Dies lässt den Schluss zu, dass die Trennbarkeit auch mit der untersuchten Population zusammenhängt (Kleickmann & Anders, 2011). Im vorliegenden Projekt ist aufgrund der Ausbildungsstrukturen von einem engen Zusammenhang von Fachwissen und fachdidaktischem Wissen/Können auszugehen, da die Probandinnen und Probanden an Pädagogischen Hochschulen ausgebildet wurden, wo die beiden Wissensbereiche kaum getrennt unterrichtet werden.

8.1.4 Validität des Befragungsinstrument

Ein wichtiges Ziel der vorliegenden Studie ist die Validierung des Befragungsinstruments. Dazu wurden vier Forschungsziele formuliert die im Folgenden zusammengefasst und diskutiert werden. Ein erstes Ziel hierzu lautet: *Das Befragungsinstrument soll inhaltsvalide sein,*

indem theoretisch sowie praktisch relevante Inhalte erfasst werden, die durch Expertinnen und Experten als solche beurteilt werden.

Inhaltsvalidität und Standards des Professionswissens

Wie bereits im Zusammenhang mit der Interpretationsobjektivität erwähnt wurde, muss diskutiert werden, welche inhaltlichen Setzungen vorgenommen wurden und wie festgelegt wurde, was überhaupt als relevantes, gültiges Wissen angesehen wird (Wiprächtiger-Geppert et al., 2015). Vorstellungen von professionellem Leseförderhandeln und dessen Wissensgrundlagen müssen expliziert werden, da nur durch die explizite Modellierung professionellen Wissens und Handelns normative Ansprüche diskutiert werden können (Scherf, 2016).

In der vorliegenden Studie wurden Items entwickelt, die sich auf die aktuelle Fachliteratur im Bereich Lesen und LRS beziehen und sich durch empirische Forschungsergebnisse abstützen lassen (Kap. 2). Dabei wurde der Anspruch verfolgt, Items zu entwickeln, die relativ unabhängig von gewissen lesedidaktischen Strömungen sind (Riegler & Wiprächtiger-Geppert, 2016). Alle Items wurden den Dimensionen des theoretischen Modells zugeordnet. Damit konnte eine inhaltliche Abdeckung des Konstrukts *Professionswissen zu Grundlagen des Leseerwerbs und Diagnostik bei Leseschwierigkeiten* umgesetzt werden. In der Hauptuntersuchung wurden 28 Items geprüft, die das theoretische Modell durch eine möglichst gleichmäßige Verteilung abdecken. Durch die Reduktion des Befragungsinstruments auf 18 Items, die akzeptable statistische Kennwerte aufwiesen, ist die inhaltliche Validität insofern weiterhin erfüllt, als jede Facette zumindest durch zwei Items repräsentiert ist (reduziertes Modell im Anhang).

Ein weiterer Bezugspunkt zur Absicherung von relevantem, validem Wissen ist der bereits weiter oben erwähnte Einbezug von Standards im Sinne normativer Kompetenzerwartungen (Kämper-van den Boogaart, 2010). So haben die Standards der Kulturreministerkonferenz (z.B. Kulturreministerkonferenz [KMK], 2014) bei der Entwicklung zahlreicher Messinstrumente zum Professionswissen im Lehrberuf Eingang gefunden (Bremerich-Vos & Dämmer, 2013; Pissarek & Schilcher, 2017; Rutsch, 2016). Diese Standards stellen einen wichtigen Konsens über zu vermittelnde Wissensinhalte in der Lehrerinnen- und Lehrerbildung dar und sind mitbestimmend dafür, welches Wissen in den Instrumenten als relevant angesehen wird (Spinath, 2019). In der vorliegenden Untersuchung wurde auf diesen Bezug allerdings verzichtet, da die formulierten Standards auf Primarstufe sehr allgemein gehalten sind und sich in einer stichwortartigen Beschreibung des vorgegebenen Wissens im Bereich Lesen erschöpfen (Kulturreministerkonferenz [KMK], 2019). Für die Weiterentwicklung des Instruments ist die

Frage nach einer kriterienorientierten Messung von Wissen/Können mit Zielvorgaben – oder eben Standards – die auch messbar sind (Grollmann & Jude, 2008) allerdings wichtig.

Validierung durch Expertinnen und Experten

Neben der theoretischen Validierung erfolgte eine weitere Abstützung der Gültigkeit von Wissensbeständen durch die Befragung von Expertinnen und Experten. Das Vorgehen hat sich in verschiedenen Studien zur Optimierung von Instrumenten bewährt (Blömeke & König, 2010; Krauss et al., 2011; Pissarek & Schilcher, 2017; Vogelsang & Reinhold, 2013). Auch im vorliegenden Projekt wurden die Items hinsichtlich ihrer praktischen und theoretischen Relevanz sowie ihrer Lösbarkeit von Expertinnen und Experten beurteilt (Kap. 5.3.3.1). Zentrales Anliegen dieses Validierungsschrittes war das Aufspüren von Inhalten und Konzeptionen, die für das Unterrichten im Bereich Lesen relevant sind. Das bedeutet, dass ein Nicht-Wissen in diesem Bereich folgenschwere praktische Konsequenzen hätte (Krebs, 2004). Ob dies tatsächlich so ist, lässt sich aber letztlich nur durch eine prädiktive Validierung belegen, d.h. über die direkte Messung der Konsequenzen wie z.B. schlechte Leseleistungen oder das Nicht-Erkennen von Leseschwierigkeiten. Die Befunde zur prädiktiven Validität von Messinstrumenten – d.h., ob die Tests in der Lage sind, Unterschiede im unterrichtlichen Handeln und Schülerinnen- und Schülerleistungen zu prognostizieren – sind jedoch noch recht dürftig (Brühwiler et al., 2017). Zu diskutieren ist, ob Normen von Expertinnen und Experten ihrem Anspruch, relevantes, berufspraktisches Wissen abzusichern, überhaupt gerecht werden. Zum einen ist fraglich, inwiefern diese Normen prädiktiv sind und mit Faktoren wie Schülerinnen- und Schülerleistungen in Einklang stehen, zum anderen, ob diese in hohem Masse an fachdidaktischer und pädagogisch-psychologischer Forschung orientierten Normen nicht fern des berufspraktischen Handelns liegen (Neuweg, 2015). Diese Frage ist allerdings etwas weniger relevant bei einem Befragungsinstrument, das *kognitives* Ausbildungswissen erfasst, wie dies in der vorliegenden Untersuchung der Fall ist. Wird allerdings der Anspruch gestellt, handlungsleitendes, kontextualisiertes Wissen zu erfassen, so steht das dem Handeln eingeschriebene implizite Wissen zunehmend quer zum Ausbildungswissen (Bromme, 1992) und lässt sich kaum über Normen von Expertinnen und Experten validieren. Damit wird eine zentrale Frage zum Validitätsanspruch des vorliegenden Befragungsinstruments aufgeworfen, nämlich die Frage nach dem Zusammenhang von ausbildungsbezogenem kognitivem Wissen und dessen Umsetzung in realen Anforderungssituationen. Damit wird der Anspruch der *praktischen Relevanz* des Wissens bzw. der Validität von Leistungsdispositionen hinsichtlich verschiedener Zielkriterien thematisiert (Brühwiler et al., 2017). Die Frage hängt dabei stark mit messmethodischen Aspekten

zusammen. Der Anspruch der kontextualisierten Erfassung von Kompetenz setzt am erweiterten Kompetenzbegriff an, bei dem Professionswissen über spezifische Situationen in sichtbares Handeln übergeht (Kap. 3.1.1). Im vorliegenden Instrument wurde diese Handlungsnähe im Rahmen der Audioitems sowie bei Items, bei denen Lehrpersonen selbst Aufgaben entwickeln mussten realisiert. Da dies (u.a. aus bereits genannten ökonomischen Gründen) nur relativ wenige Items betrifft, bewegt sich das Befragungsinstrument vergleichsweise weit weg vom Anspruch der Kompetenzmessung in realen Situationen. Kontextsensitivität im Sinne einer sichtbaren Transformation von Wissen/Können in Handlungen bleibt ein anzustrebendes Ziel, dem aber immer auch Grenzen gesetzt sind, sobald das praktische Können außerhalb der Realsituation (z.B. über unterrichtsnahe Fallvignetten) erfasst wird (Neuweg, 2014). Empirische Befunde zeigen, dass videobasierte Vignetten gegenüber textbasierten Vignetten nicht notwendigerweise vorteilhafter sind, um unterrichtsrelevantes Wissen zu erfassen (Brühwiler et al., 2017). Der Befund, dass eine mittels Textvignetten erfasste höhere Wissensausprägung nicht zwingend mit einer höheren Unterrichtsqualität in der Realsituation korrespondiert, findet sich in einer Studie mit Physiklehrpersonen (Vogelsang & Reinhold, 2013). Offensichtlich werden mit stärker kontextualisierten Instrumenten die Validitätsprobleme nicht unbedingt gelöst.

Künftig wäre es wichtig, Untersuchungen zur prädiktiven Validität verstärkt auf den Lernzuwachs der Schülerinnen und Schüler zu richten und zu erheben, ob Instrumente nicht nur unterrichtsrelevantes, sondern auch *lernrelevantes* Wissen erfassen (Brühwiler et al., 2017). Im Rahmen der COACTIV-Studie scheint dies gelungen zu sein, für andere Fächer und Domänen steht der Nachweis noch aus. Effekte einzelner Wissensbereiche und Wissensfacetten hinsichtlich verschiedener Kriteriumsvariablen könnten aufschlussreich sein, „um mehr über die Wege zu erfahren, über die Lehrerwissen seine Wirksamkeit auf das Lernen der Schüler entfaltet“ (Neuweg, 2014, S. 603). Dieser nächste Schritt einer kriterienorientierten Validierung steht für das vorliegende Instrument aus und müsste im Rahmen einer Weiterentwicklung verfolgt werden. Es wäre zu zeigen, dass auch für den Bereich Lesen gilt, was für den Bereich Mathematik bereits festgestellt wurde, nämlich dass sich kognitive Wissensbestände auf unterrichtliches Handeln und schliesslich auf die Schülerinnen- und Schülerleistungen auswirken (Kap. 3.4.3).

8.1.5 Validität des Befragungsinstruments im Rahmen von Gruppenvergleichen

Eine weitere Anforderung an Messinstrumente besteht darin, dass in den zu vergleichenden Gruppen dasselbe latente Konstrukt gemessen wird und damit Messinvarianz vorliegt. Das entsprechende Forschungsanliegen lautet: *Mit dem Befragungsinstrument soll in verschiedenen, nach Berufsbiografie gebildeten Substichproben, dasselbe Konstrukt erfasst werden.*

Ergebnisse der Differenziellen Itemfunktionsanalyse

In der vorliegenden Studie hat sich gezeigt, dass der Anspruch, über alle Substichproben dasselbe Konstrukt zu messen, gemäß Ergebnissen der DIF-Analysen nicht erfüllt wurde (Kap. 7.2.3). Damit sind Einschränkungen in der Konstruktvalidität nachgewiesen. Der paarweise Vergleich der drei Substichproben zeigte, dass lediglich sechs Items durchgängig über alle Substichproben hinweg gleich gelöst wurden. Bei allen anderen Items lagen Abweichungen im Differenzbetrag der Logit-Werte von ≥ 0.43 und/oder eine signifikante Abweichung von $t > 1.96$ vor (Kap. 7.2.3.2, Tabelle 23). Die größten Abweichungen zeigten sich im Vergleich zwischen Substichproben $\text{Primar}_{\text{mitBE}}$ vs. SHP und Substichproben $\text{Primar}_{\text{ohneBE}}$ vs. SHP. Daraus lässt sich umgekehrt ableiten, dass die Substichproben $\text{Primar}_{\text{ohneBE}}$ und $\text{Primar}_{\text{mitBE}}$ zumindest ähnlichere Konzepte zum Lösen der Aufgaben verwendeten und dass diesbezüglich deutlichere Unterschiede im Vergleich zu SHP vorlagen.

Da Verstehens- und Erinnerungsprozesse über verschiedene Ausbildungsphasen hinweg das Lösen von Items beeinflusst, sind Gruppenunterschiede grundsätzlich zu erwarten. Hier stellt sich die Frage zum Umgang mit DIF. Würde man den Empfehlungen folgen und von DIF betroffene Items für weitere Analysen ausschließen (Borsboom, 2006; Tennant & Pallant, 2007), so müssten Einschränkungen hinsichtlich der inhaltlichen Validität in Kauf genommen werden. Mit einer Reduktion von Items besteht das Risiko, dass die Items das Wissensspektrum nicht angemessen einfangen und dass der Einsatz des Instruments für die gewollt heterogene Gruppe in Frage gestellt ist (Schulte et al., 2013). In der vorliegenden Studie wurden *drei* Substichproben verglichen und damit eine äußerst heterogene Gruppe getestet. Da die Schnittmenge von Items, die in unterschiedlich zusammengesetzten Substichproben Messinvarianz aufweisen tendenziell kleiner ist, ergibt sich hier ein forschungsmethodisches Dilemma (ebd.). Dieses zeigt sich darin, dass eine hohe Heterogenität der Gruppe auch zu mehr DIF führt, dass jedoch auf der anderen Seite nur eine genügend hohe Anzahl und Variation von Items das heterogene Wissensspektrum angemessen erfasst. „Für inhaltliche Interpretationen stößt das methodische Konzept der statistischen Überprüfung von Messinvarianz an seine Grenzen“ (ebd., S. 114). Bei Verletzung der Messinvarianz eines Items oder Instruments muss die Diskussion daher auch auf theoretischer Ebene erfolgen (ebd.). Da der Verwendungszweck des vorliegenden Befragungsinstrumentes ja gerade darin besteht, Stichproben mit unterschiedlichem berufsbiografischem Hintergrund und ausbildungsbezogenem Wissen zu vergleichen, wurde hier anstelle eines Ausschlusses betreffender Items ein interpretativer Umgang gewählt. Die DIF wird nach diesem Ansatz als wertvolle Informationsquelle für weiterführende inhaltliche Analysen

genutzt, und es kann nach möglichen Gründen für unterschiedliche Antwortmuster gesucht werden (Schulte et al., 2013; Zumbo, 2007).

Aus den vorliegenden Daten wurden drei mögliche „Wissenskomplexe“ herausgefiltert, um gruppenspezifische Antwortmuster zu erklären (Kap. 7.2.4.4). Zunächst lassen sich Items zu *spezifisch heilpädagogischem Wissen* identifizierten. Es liegt auf der Hand, dass SHP aufgrund ihres vertieften Wissens mit einigen Inhalten vertrauter sind als die anderen Substichproben. Ein weiterer Erklärungskomplex bildet sich um eine Gruppe von Items, bei deren Lösung das *Berufserfahrungswissen* eine Rolle spielen könnte, die also von Substichprobe *Primar_{mitBE}* überproportional besser gelöst wurden. Einen dritten Komplex bilden Items, bei denen es um Aspekte der *Leseflüssigkeit* geht und die von Substichprobe *Primar_{ohneBE}* erwartungswidrig besser gelöst wurden. Eine plausible Erklärung für dieses unterschiedliche Lösungsverhalten fällt allerdings schwer. Im Rahmen weiterführender Studien wäre zu klären, ob mit dem Begriff Leseflüssigkeit möglicherweise bestimmte Assoziationen verbunden sind, die sich in den Gruppen unterscheiden. Alle Interpretationen sind als Arbeitshypothesen zu verstehen, zu deren Überprüfung qualitative Analysen über ablaufende kognitive Prozesse beim Lösen der Items durchgeführt werden müssten.

Da der Nachweis von Messinvarianz für den Vergleich von Gruppenmittelwerten wünschenswert ist, wurde im vorliegenden Fall die Robustheit messinvarianter Skalen durch einen Modellvergleich mit drei unterschiedlich gebildeten Skalen geprüft (siehe Vorgehen bei HarmoS: Ramseier, 2008). In den verglichenen Skalen wurde die Messinvarianz durch das Splitten von Items operationalisiert (Kap. 7.2.3.4). Die Ergebnisse zeigten, dass sich die Mittelwerte der drei Substichproben in allen drei Skalen (vollständig gesplittete, teilweise gesplittete und nicht gesplittete Items) in derselben Richtung unterscheiden (Kap. 7.2.3.4, Tabelle 24). Dies kann als Hinweis auf eine ausreichende Robustheit der Basis-Skala – mit nicht gesplitteten Items – gewertet werden. Unter dem Gesichtspunkt einer besseren inhaltlichen Abdeckung empfiehlt es sich daher, die Basis-Skala für weitere Analysen zu verwenden, allerdings mit nachdrücklichem Verweis darauf, dass diese Skala stichprobenabhängig Unterschiedliches misst.

Ausbildungssensitivität

Ein weiterer zentraler Anspruch lag darin, ein Befragungsinstrument zu entwickeln, das ausbildungssensitiv misst. Das entsprechende Ziel lautet: *Mit dem Befragungsinstrument sollen Wissensunterschiede zwischen Lehrpersonen mit verschiedener Berufsbiografie (Ausbildung/Berufserfahrung) sensitiv erfasst werden.*

Ergebnisse der einfaktoriellen ANOVA und der Einzelvergleich mittels Games-Howell-Test zeigten, dass das entwickelte Befragungsinstrument ausbildungs- und berufserfahrungsbezogene Gruppenunterschiede zu erfassen vermag und damit ein wichtiges Validitätskriterium erfüllt ist (Kap. 7.2.5.1). Der ermittelte Haupteffekt aus der ANOVA ($\eta^2 = 0.399$) verweist auf eine hohe Varianzaufklärung (Döring & Bortz, 2016) des gemessenen Professionswissens durch den Faktor Gruppenzugehörigkeit. Verglichen wurden wiederum die drei Substichproben $\text{Primar}_{\text{ohneBE}}$, $\text{Primar}_{\text{mitBE}}$ und SHP. Der querschnittliche Anstieg des Professionswissens im Bereich Lesen zeigte sich in erwarteter Richtung von Substichprobe $\text{Primar}_{\text{ohneBE}}$ über Substichprobe $\text{Primar}_{\text{mitBE}}$ zum höchsten Mittelwert der Substichprobe SHP (Kap. 7.2.5). Die Effektstärken zwischen den Gruppen können ebenfalls als bedeutsam eingestuft werden und erreichen die Werte $d = 0.55$ (Substichprobe $\text{Primar}_{\text{ohneBE}}$ vs. $\text{Primar}_{\text{mitBE}}$) und $d = 1.29$ (Substichprobe $\text{Primar}_{\text{mitBE}}$ vs. SHP) (Kap. 7.2.5.2, Tabelle 30).

Dieses Ergebnis lässt sich einerseits vor dem theoretischen Hintergrund der Expertiseforschung einordnen, andererseits vor dem Hintergrund empirischer Studien, in denen die Wissensentwicklung über verschiedene Phasen der Ausbildung untersucht wurde. Die vorliegenden Ergebnisse stützen die These des Expertiseansatzes grundsätzlich dahingehend, dass sich kognitive Wissensbestände von Novizinnen und Novizen auf dem Weg zu Expertinnen und Experten entwickeln und über verschiedene Erwerbsphasen im Rahmen der Ausbildung mitbestimmt werden (Kap. 3.1.2). Allerdings muss hier erwähnt werden, dass der Wissens- und Könnenszuwachs lediglich querschnittlich und nicht längsschnittlich untersucht wurde und deshalb vorsichtig interpretiert werden muss (Blömeke, 2013). Andere Studien, in denen querschnittliche Leistungsvergleiche von Studierenden im Studienverlauf (vor allem auf der Sekundarstufe I) untersucht wurden, zeigen eher widersprüchliche Ergebnisse hinsichtlich der Entwicklung von Fachwissen und fachdidaktischem Wissen. Während die meisten Untersuchungen relativ einheitlich zeigen, dass sich das fachdidaktische Wissen im Rahmen der Ausbildung entwickelt (Baer et al., 2007; Blömeke et al., 2013a; Klusmann, 2011; Lindl & Krauss, 2017; Schmidt et al., 2011), scheint dies für das Fachwissen eher nicht der Fall zu sein (Baer et al., 2007; Blömeke et al., 2013a). In Bezug auf die (nicht-)gemessenen Leistungszuwächse im Fachwissen der erwähnten Studien wurde interpretiert, dass dies u.a. auch mit dem Testzeitpunkt zusammenhänge bzw. mit der zeitlichen Nähe der während der Ausbildung vermittelten Inhalte und deren Erfassung durch das Befragungsinstrument (Bremerich-Vos & Dämmer, 2013). Studierende entfernen sich im Laufe ihrer Ausbildung von den auf basales Wissen zielende Items (Bremerich-Vos & Dämmer, 2013).

In der vorliegenden Arbeit wurden relativ weit auseinanderliegende Ausbildungsphasen bzw. Messzeitpunkte gewählt (erstes und fünftes Semester). Zudem handelte es sich bei der Ausbildung in Schulischer Heilpädagogik um eine *Zusatzausbildung* mit teilweise deutlich anderem Fokus als in der Grundausbildung, so dass die eindeutigen Wissensunterschiede auch damit zusammenhängen dürften. Da die strukturelle Unterscheidung von Fachwissen und fachdidaktischem Wissen in der vorliegenden Studie empirisch nicht belastbar war, können keine Aussagen zu unterschiedlichen Entwicklungsverläufen der beiden Dimensionen Fachwissen und fachdidaktisches Wissen gemacht werden.

Im Weiteren ist der Einfluss der Berufserfahrung zu diskutieren, die sich in der vorliegenden Untersuchung beim Vergleich der Substichprobe *Primar_{ohneBE}* vs. *Primar_{mitBE}* ($d = 0.55$) nur mäßig im gemessenen Professionswissen niederschlägt. Die widersprüchlichen Befunde zum Einfluss der Berufserfahrung anderer Studien sind in Kap. 3.4.2 dargelegt. Sie lassen sich so zusammenfassen, dass der Effekt der Berufserfahrung auf das Professionswissen in vielen Studien ausblieb (Brunner et al., 2006; Jandl, 2016; Lindl & Krauss, 2017; Pissarek & Schilcher, 2017), während sich die Lehrerfahrung im Leseunterricht z.B. in der Studie von Rutsch (2016) auf das *lesedidaktische Wissen* nachweisen ließ. Allerdings ist der Einfluss der Lehrerfahrung auf den latenten Faktor *lesedidaktisches Wissen* auch hier als gering einzuschätzen (Rutsch, 2016; Rutsch et al., 2018b). Auch wenn nicht bezweifelt wird, dass mit zunehmender Berufserfahrung für den Lehralltag wichtige Routinen und Automatisierungen erworben werden, so bleibt offen, ob die Berufserfahrung tatsächlich weniger bedeutsam ist oder ob das Instrument selbst den Anspruch der berufserfahrungsbezogenen Veränderungsmessung nicht einlösen kann (Krauss & Bruckmaier, 2014). Im vorliegenden Fall reagiert das Instrument zwar auf berufserfahrungsbezogene Gruppenunterschiede. In welchem Masse „*Erfahrungswissen*“ tatsächlich erfasst wird, ist allerdings zu diskutieren. Die genaue Beschreibung dessen, was in unserem Alltagsverständnis als Erfahrungswissen bezeichnet wird, scheint eine besondere Herausforderung zu sein (Combe & Kolbe, 2008). Es zeigt sich nämlich, dass es sich beim Einsatz von Erfahrungswissen nicht einfach um „eine schlichte Wissensanwendung, bei der das Handeln aus einem ‘handlungssteuernden Wissen‘ gleichsam abgeleitet wird“, handelt (Combe & Kolbe, 2008, S. 862). Erfahrungswissen schlägt sich demnach in anderen Strukturen nieder, die meist als Interpretationsmuster, Handlungsmuster oder Routinen beschrieben werden (Combe & Kolbe, 2008; Neuweg, 2014). Diese Art von Wissen und Können wird mit dem entwickelten Instrument nicht erfasst, so dass sich das Erfahrungswissen auch deshalb bei der Gruppe der *Primar_{mitBE}* weniger deutlich niederschlagen dürfte.

Abschließend soll noch ein Blick auf die Gruppe der SHP geworfen werden. Studierende am Ende der Ausbildung der Schulischen Heilpädagogik weisen in der vorliegenden Studie mit Abstand ein höheres Professionswissen auf als Primarlehrpersonen mit Berufserfahrung. Dieses Ergebnis schließt an dasjenige der Studie von Jandl (2016) an, die zeigt, dass Studierende der Schulischen Heilpädagogik am Ende ihrer Ausbildung ein signifikant höheres mathematikspezifisches Professionswissen aufweisen als Lehrpersonen mit anderer oder fehlender (heil-) pädagogischer Ausbildung und auch als berufserfahrene SHP mit bereits abgeschlossener Berufsausbildung. Auch die Studie von Schmidt und Schabmann (2016) zeigt, dass Lehrpersonen der Sonderpädagogik über ein größeres Grundwissen im Bereich Lesen verfügen als die Primar- und Sekundarlehrpersonen. In eine andere Richtung weisen hingegen die Ergebnisse der Studie von Rutsch (2016), in der im Bereich Lesedidaktik kein Wissensunterschied zwischen Primarlehrpersonen und Lehrpersonen der Sonderpädagogik nachgewiesen werden konnte. In vergleichbarer Richtung weist auch die Studie von Hintz und Grünke (2009), in der kein Wissensvorsprung sonderpädagogischer Lehrpersonen hinsichtlich der Förderung des Schriftspracherwerbs bei Kindern mit gravierenden Leistungsproblemen festgestellt wurde. Hier wäre danach zu fragen, inwiefern die widersprüchlichen Ergebnisse mit den Instrumenten selbst zusammenhängen. Denkbar wäre, dass fehlende Unterschiede im Professionswissen der Berufsgruppen möglicherweise durch die Erfassung mit nicht-ausbildungssensitiven Instrumenten zustande kommen und dass damit ein Validitätsproblem vorliegt. Dieses Validitätsproblem bei Untersuchungen zeigt sich deutlich, wenn die Validität eines Erhebungsinstruments und der eigentliche Prozess der Validierung nicht voneinander getrennt betrachtet werden (Messick, 1995). In der vorliegenden Arbeit wurde der Entwicklungsprozess des Befragungsinstruments strikte von einer eigentlichen Untersuchung getrennt, wodurch ein Validitätsproblem entschärft werden konnte. Hiermit liegt ein Instrument vor, das Ausbildungsunterschiede nachweislich erfassen kann. Dies ist ein deutlicher Vorteil des vorliegenden Befragungsinstruments gegenüber Instrumenten, in denen dies nicht vorgängig getestet wurde.

Kontrastgruppenvergleich

Im Rahmen des Validierungsprozesses stellt sich die Frage, ob es sich beim erfassten Konstrukt tatsächlich um *Professionswissen* handelt, das sich im Rahmen der Ausbildung entwickelt, oder ob eher allgemeine Wissensbestände abgefragt wurden. Entsprechend wurde folgendes Ziel geprüft: *Mit dem Befragungsinstrument soll lesespezifisches Professionswissen erfasst werden, das sich deutlich von allgemeinen Wissensbeständen von Laien abgrenzt.*

Beim Einzelvergleich der genannten Substichproben wurde in einem weiteren Schritt die Kontraststichprobe der Laien einbezogen. Erwartet wurde, dass sich diese Substichprobe bezüglich Professionswissen deutlich von den anderen Gruppen abhebt. Zusammengesetzt war die Gruppe aus Studierenden, die am Anfang der Ausbildung zur Grundschullehrperson oder der Ausbildung in Schulischer Heilpädagogik standen und kein Fachwissen zu Themen wie Schriftspracherwerb und Umgang mit Leseschwierigkeiten hatten. Die Ergebnisse der Mittelwertvergleiche zeigten, dass sich Laien erwartungsgemäß von den übrigen drei Substichproben unterschieden (Kap.7.2.5.1). Der Wissensunterschied zwischen der Laien-Kontraststichprobe und der Substichprobe Primar_{ohneBE} drückte sich in der Hauptuntersuchung in einer mittleren Effektstärke ($d = 0.60$) aus (Döring & Bortz, 2016, S. 820). Damit kann dem vorliegenden Befragungsinstrument grundsätzlich bescheinigt werden, professionsspezifisches Wissen abzubilden.

Zu diskutieren ist, ob im Hinblick auf die Validität des Befragungsinstruments ein noch deutlicherer Unterschied zur Kontraststichprobe erstrebenswert wäre, um zu belegen, dass nicht auch beiläufig erworbenes Wissen oder gute Ratefähigkeiten erfasst wurden. Rateeffekte sind bei gebundenen Antwortformaten immer zu erwarten und hängen mit der Plausibilität von Distraktoren zusammen (Bühner, 2011). Eine Herausforderung stellt sich dann, wenn Wissen ohne die Hürde einer spezifischen Fachsprache erfasst werden soll, was unter Umständen auf Kosten der fachlichen Spezifik der Aufgaben gehen kann (Wiprächtiger-Geppert et al., 2015). Im vorliegenden Instrument wurde auf fachspezifische Formulierungen wenn möglich verzichtet. Damit steigt das Risiko, allgemeines bzw. nach Plausibilitätskriterien organisiertes Wissen und Können abzufragen, das auch bei Laien vorliegen kann. Die Ergebnisse der Distraktorenanalyse (Kap. 7.2.2) weisen bei einigen Items auf die Anwendung von Ratestrategien oder auf eine Plausibilitätsauswahl von Antworten hin.

Eine weitere Überlegung hinsichtlich der Validierung durch eine Laienstichprobe betrifft die Stichprobenzusammensetzung. Es hat sich nämlich gezeigt, dass der Wissensunterschied in der Voruntersuchung zwischen der Laien-Kontraststichprobe und der Substichprobe Primar_{ohneBE} wesentlich größer war ($d = 1.22$). Dies lässt sich möglicherweise mit der unterschiedlichen

Stichprobenzusammensetzung der Laien in der Vor- und in der Hauptuntersuchung erklären, da in der Voruntersuchung Personen aus dem privaten Umfeld befragt wurden, die nicht so nahe am schulischen Kontext waren wie die Stichprobe der Hauptuntersuchung. Die Schlussfolgerung liegt nahe, dass sowohl die Ausgestaltung des Instruments als auch die Stichprobenzusammensetzung einen Einfluss auf die diskriminante Validität haben. So zeigen die Ergebnisse von Rutsch (2016) z.B. keinen deutlichen Effekt zwischen der Untersuchungsgruppe (Deutschlehramtsstudierende) und den Kontrollgruppen (fachfremdes Lehramt, Psychologiestudierende, Germanistikstudierende), wohingegen die ähnlich angelegten kontrastierenden Analysen der COACTIV-Studie Hinweise auf Professionsspezifität bzw. deren Erfassbarkeit durch das entsprechende Befragungsinstrument geben (Krauss et al., 2011).

8.2 Limitation des Befragungsinstruments

8.2.1 Theoretische Aspekte

Das vorliegende Befragungsinstrument hat zwei wichtige Wissensbereiche im Blick, die für im Beruf stehende und angehende Lehrpersonen auf Primarstufe relevant sind (Wehr & Hartmann, 2005). Es fokussiert auf das Grundlagenwissen und das (diagnostische) Wissen zu „Leseschwierigkeiten“. Der Begriff der Leseschwierigkeit wird, wie in Kap. 2.2.1 dargelegt, bewusst weit gefasst, es wird also nicht hochspezialisiertes Wissen über Kinder mit einer klassifizierten LRS abgefragt. Eine deutliche Abgrenzung zwischen „Wissen über LRS“ und „Wissen über Leseschwierigkeiten“ wurde aber nicht vollzogen, da bezweifelt wird, dass eine solche Abgrenzung überhaupt sinnvoll ist. In vielen diagnostischen Verfahren wird davon ausgegangen, dass mit gravierenden Folgen für betreffende Kinder zu rechnen ist, auch wenn sich die Normabweichungen als geringer erweisen als die Werte, die in der ICD-10 und DSM-5 (Kap. 2.2.2, Falkai & Wittchen, 2015; Schulte-Körne, 2014b) angegeben werden (Klicpera et al., 2017).

Eine inhaltliche Limitation des vorliegenden Befragungsinstruments betrifft den fehlenden Einbezug des *Förderwissens*, das aus ökonomischen Gründen weitgehend ausgeklammert wurde. Dieses Wissen ist unbestritten relevant und muss in zukünftig zu entwickelnden Instrumenten in den Blick genommen werden.

Eine weitere Einschränkung bezieht sich auf die Auswahl und Beständigkeit des erfassten Wissens und Könnens. Alle Items sind theoretisch und empirisch abgestützt und decken das Modell des lesebezogenen Wissens/Könnens (Kap. 4.1) im Sinne *möglicher* Operationalisierungen ab. Damit wird gleichzeitig klar, dass eine Auswahl getroffen wurde und dass Items stellvertretend für den ganzen Wissensbereich stehen (Rutsch, 2016). Der empirische Nachweis, dass sich die

exemplarisch erfragten Inhalte tatsächlich auf unterrichtliches Handeln oder letztlich auch auf die Leseleistungen der Schülerinnen und Schüler niederschlagen, wäre in Folgestudien noch zu erbringen. Die aktuelle Forschungslage über derartige Zusammenhänge im Bereich Lesen ist derzeit noch zu spärlich, um ein abschließendes Fazit ziehen zu können (Kap. 3.4.1). Eine inhaltliche Einschränkung besteht auch hinsichtlich des fehlenden Einbezugs von Themen wie Lesemotivation (Guthrie & Wigfield, 2000; McElvany et al., 2008), Lesesozialisation (Hurrelmann, 2004), besondere Situation von Kindern mit Deutsch als Zweitsprache (Schründer-Lenzen & Merkens, 2006) sowie bezüglich des Einflusses digitaler Medien auf das Lesen (Bertschi-Kaufmann, 2007). Diese Themen wurden im vorliegenden Instrument nicht berücksichtigt.

Limitationen ergeben sich auch aus der formalen und theoretischen Modellierung des zugrunde gelegten Modells des lesebezogenen Wissens/Könnens. Da sich die postulierte Struktur in den Daten nicht widerspiegelt, stellt sich die Frage, ob die Übertragung von Fachwissen und fachdidaktischem Können auf den Bereich Lesen nicht gelingt, weil sich der Wissensbereich Lesen strukturell von anderen Fächern (z.B. Mathematik) unterscheidet (Phelps & Schilling, 2004) oder ob die zwei Bereiche durch die Items nicht hinreichend trennscharf operationalisiert waren. Für das vorliegende Instrument gilt deshalb, was Hill et al. (2008) nach ihrem gescheiterten Versuch, ihr Modell empirisch zu belegen, festhalten: „In sum, our work suggests that the conceptualization of the domain is far from straightforward. In our case, we attempted to build the ship while sailing it – by writing items to help push our conceptualization and definition forward“ (S. 396). Möchte man verschiedene Qualitäten von Wissen erfassen, um mögliche Zusammenhänge zwischen diesen und dem unterrichtlichen Handeln oder den Schülerinnen- und Schülerleistungen zu untersuchen, müssen diesbezüglich weitere Anstrengungen unternommen werden (Lessing-Sattari & Wieser, 2018). Festzuhalten ist jedoch, dass das vorliegende Instrument seinen Verwendungszweck auch ohne die empirische Trennbarkeit von Fachwissen und fachdidaktischem Können bzw. Grundlagenwissen und diagnostischem Wissen erfüllt.

Gravierender hingegen sind die Einschränkungen der inhaltlichen Validität aufgrund des Ausschlusses von Items, die psychometrischen Kriterien nicht standhalten konnten (Kap. 7.3). Durch die Reduktion der ursprünglichen Skala mit 28 Items auf eine Skala mit 18 Items ist die Abdeckung des theoretischen Modells zum lesebezogenen Wissen und Können eingeschränkt, und es müssten in einem weiteren Schritt neue Items entwickelt bzw. bestehende überarbeitet werden.

Ein weiterer Aspekt, der zu bedenken ist, hängt mit der Eingrenzung auf kognitiv abrufbare Wissensbestände und mit der mehrheitlich kontextfernen Operationalisierung von Wissen und

Können zusammen. Im vorliegenden Befragungsinstrument wurde nicht *Kompetenz* erfasst, sondern *Professionswissen* und *Professionskönnen* (Kap. 3.1.1). Dimensionen wie motivationale Orientierung, Überzeugungen oder Selbstregulation wurden nicht berücksichtigt (Kap. 3.2.2). Inwiefern diese und weitere Variablen wie Studienort, Geschlecht oder individuelle Faktoren das Professionswissen beeinflussen (Blömeke, Buchholtz & Bremerich-Vos, 2013b), kann somit nicht beantwortet werden und wäre im Rahmen weiterführender Untersuchungen zu prüfen. Die eher kontextferne Erfassung kognitiv abrufbaren Wissens und Könnens durch weitgehend geschlossene Formate hat zudem zur Folge, dass die weitaus komplexeren Handlungsvollzüge im Unterricht möglicherweise nicht direkt an diesem Wissen anschließen (Blömeke et al., 2015b). Für eine spätere Verwendung des Instrumentes müssten die verschiedenen Unsicherheiten, die auf dem Weg zwischen Wissen, Können, Handeln und Schülerinnen- und Schülerleistung liegen, weiter analysiert werden (Brühwiler et al., 2017; Scherf, 2016). Trotz der Unsicherheiten ist jedoch davon auszugehen, dass das Wissen das Handeln – zumindest teilweise – mitsteuert. Anders würden sich die Effekte des Fachwissens und des fachdidaktischen Wissens auf Schülerinnen- und Schülerleistungen schwer erklären lassen (Kunter et al., 2011a). Studierende, die in der hier vorgelegten Untersuchung über hohes Wissen und Können verfügen, besitzen damit die fachlichen Voraussetzungen für professionelles Handeln; dies ist eine notwendige, aber nicht hinlängliche Bedingung. Andererseits ist nicht ganz auszuschließen, dass Studierende mit geringem Wissen trotzdem guten (Lese-)Unterricht gestalten und ihre Lücken im Fachwissen durch andere pädagogische Fähigkeiten kompensieren (Neuweg, 2014). Solches Wissen ist aber eher problematisch, denn in diesem Fall könnte es sich möglicherweise auch um zufälliges Wissen handeln. Um die Fragen nach den Wirkzusammenhängen zu klären, wären Extremgruppenvergleiche wie in der Studie von Piasta et al. (2009) aufschlussreich. Zudem müssten weitere Faktoren, die das Lernen steuern, einbezogen werden (Angebots-Nutzungsmodell: Helmke, 2012, S. 71). Letztlich lassen sich die langlaufenden Wirkungsketten, in denen das professionelle Wissen und Können implementiert ist, insgesamt nie vollständig überprüfen. Es werden „immer nur bestimmte Ausschnitte aus dem gesamten Kosmos an Voraussetzungen, Einflüssen und Auswirkungen“ (Terhart, 2012, S. 8) erfasst.

8.2.2 Methodische Aspekte

Auch in methodischer Hinsicht sind Limitationen des Befragungsinstruments sowie Grenzen von dessen Aussagekraft zu verzeichnen. Eine erste Einschränkung betrifft die Stichprobensammensetzung, mit der das Befragungsinstrument entwickelt wurde. Aus forschungsökonomischen Gründen wurden angehende und ausgebildete Studierende befragt und nicht in der Berufspraxis stehende Lehrpersonen, was die Aussagekraft einschränkt (Wittmann, 2017). Für

die spätere Anwendung bleibt somit offen, ob das Instrument auch für die Stichprobe von Lehrpersonen, die in der Praxis stehen, valide Ergebnisse liefert. Eine weitere Einschränkung bezieht sich darauf, dass im Rahmen des Validierungsprozesses ein querschnittlicher Vergleich des Professionswissens über verschiedene Kohorten gemacht wurde. Die Daten sprechen für die Ausbildungssensitivität des Instruments (Kap. 7.2.5); ob es aber in der Lage ist, Leistungsentwicklungen über die *Zeit* in verschiedenen Ausbildungsphasen zu messen, müsste in echten Längsschnittstudien geprüft werden (Rutsch, 2016; Terhart, 2012).

Eine weitere Kritik bezieht sich auf die fehlende Kontrolle alternativer Erklärungen für das Lösen der Items. Kognitive Fähigkeiten und motivationale Aspekte sowie weitere mögliche Prädiktoren, die das Lösen der Aufgaben beeinflusst haben könnten, wurden nicht erfasst. Für die Weiterentwicklung wären hier Anstrengungen zu unternehmen und mehrebenenanalytische Auswertungen notwendig (Lindl & Krauss, 2017). Im Weiteren muss auch das Problem der Ratewahrscheinlichkeit genannt werden, die bei geschlossenen Formaten immer zu erwarten ist (Kap. 5.3.2). Einige Items konnten möglicherweise aufgrund von Plausibilitätsüberlegungen auch von Laien gelöst werden, wodurch die Validität der betreffenden Items in Frage gestellt ist (Kap. 6.3.2). Im Rahmen der Weiterentwicklung des Befragungsinstruments wäre dem Lösen durch Raten bzw. Ausschlussverfahren, durch ein umfassenderes kognitives Pretesting an einer deutlich größeren Stichprobe zu begegnen.

Einschränkungen gelten auch hinsichtlich der teilweise mangelhaften psychometrischen Anforderungen – insbesondere in der Substichprobe *Primar_{mitBE}* (Kap. 7.2.1). Hier ist mit deutlichen Unsicherheiten der Ergebnisse bezüglich Reliabilität, Varianz und Validität zu rechnen. Offensichtlich produziert das Befragungsinstrument in seiner aktuellen Form für diese Zielgruppe unbefriedigende Ergebnisse und müsste vor einer weiteren Verwendung überarbeitet werden. Insbesondere müssten Items überarbeitet und neu entwickelt werden, damit berufserfahrungsbezogenes Wissen und Können besser einfangen werden könnten. Von einem negativen Einfluss der Heterogenität der verwendeten Formate ist dagegen eher nicht auszugehen, da sich dieser Faktor sonst bei allen Substichproben gleichermaßen hätte auswirken müssen. Weiter stellen die vorgefundene DIF in den drei untersuchten Substichproben (Kap. 7.2.3) sowie die eher geringe faktorenanalytische Varianzaufklärung (Kap. 7.2.4.3) ein ernstzunehmendes Problem dar. In diesem Sinne ist das vorliegende Befragungsinstrument – außer für die Stichprobe der SHP, wo zufriedenstellende psychometrische Werte vorliegen – als Entwurf zu verstehen. Neben Überarbeitungen, die auf die Behebung der vorgenannten Mängel abzielen müssten, wären weitere Validitätsprüfungen notwendig, um die Qualität des Befragungsinstruments zu beurteilen. Auf Ebene der Konstruktvalidität wären Maßnahmen zur Überprüfung der

konvergenten und diskriminanten Validität entscheidend (Bühner, 2011). Aufschlussreich wäre beispielsweise der Vergleich mit Verfahren mit ähnlichem Gültigkeitsbereich (siehe Skalen Rutsch, 2016; Schmidt & Schabmann, 2016). Daneben könnte eine kriterienorientierte Validierung (z.B. Beobachtung des Leseunterrichts oder Leseleistungen der Schülerinnen und Schüler) Aufschluss darüber geben, ob das erfragte Wissen wirklich unterrichts- und/oder leistungsrelevant ist. Allerdings sind einem solchen Vorhaben aufgrund der Komplexität von Unterricht und weiterer Einflüssen, die auf das Lernen wirken, auch Grenzen gesetzt (Terhart, 2012).

Eine letzte Einschränkung betrifft die fehlende *norm-* und *kriterienorientierte* Skalierung des Befragungsinstruments (Eid & Schmidt, 2014). Es wurde darauf verzichtet, eine Interpretationsnorm (z.B. *hohes, mittleres, geringes* Professionswissen) zu definieren. Eine normorientierte Skala richtet sich an der Normalverteilung der Fähigkeiten aus und lässt eine Einordnung der Ergebnisse in ein Bezugssystem auf individueller Ebene zu (ebd.). Im Rasch-Modell wird von einer Normalverteilung der latenten Dimension ausgegangen (Eid & Schmidt, 2014; Rost, 2004). In einem weiterführenden Schritt wäre eine modellgeleitete, empirische Identifikation von Fähigkeitsstufen auf Grundlage des Rasch-Modells möglich, indem die Skala in Abschnitte unterteilt würde (Wendt & Bos, 2011). Der Vorteil des Rasch-Modells (gemeinsame Skala für Itemschwierigkeit und Personenfähigkeit) könnte dahingehend genutzt werden, dass post-hoc eine kriterienorientierte inhaltliche Beschreibung der Stufen durch entsprechende Aufgaben erfolgen würde (Hartig, 2009; Wendt & Bos, 2011). Eine kriterienorientierte Beurteilung im Sinne von „*genügendes Wissen*“ und „*ungenügendes Wissen*“ wäre allerdings auch dann auf der Grundlage theoretischer Überlegungen zu begründen (Eid & Schmidt, 2014) oder müsste in einem nächsten Schritt aus einer diskursiven Auseinandersetzung mit Expertinnen und Experten entwickelt werden.

8.3 Ertrag für Forschung und Praxis

Das vorliegende Projekt schließt eine Forschungslücke, denn bis anhin mangelt es weitgehend an Messinstrumenten, die das kognitiv abrufbare Wissen und Können von Lehrpersonen in der Domäne Lesen objektiv, reliabel und valide messen (Kap. 1). Durch den Einbezug der Stichprobe der SHP und die Fokussierung auf die Leseentwicklung im Anfangsunterricht sowie die Erweiterung auf das Wissen über Kinder mit Leseschwierigkeiten, wurde das Anwendungsfeld des vorliegenden Befragungsinstruments neu ausgerichtet.

Mit der Fokussierung auf den Bereich *Lesen* wird ein Lernbereich berührt, dessen elementare Bedeutung angesichts der fatalen Folgen ungenügender Leseleistungen unbestritten ist (Esser & Schmidt, 1993; Naumann et al., 2010; Nix, 2010; Sturm, 2012; Tacke, 2008; Wember, 1999;

Wiesner & Schneider, 2011). Die dargelegten Ergebnisse unterstreichen zudem die Bedeutung der Lehrpersonen, denen als Mitakteurinnen und Mitakteure beim Aufbau der Lesekompetenz und dem Erkennen von Leseschwierigkeiten eine zentrale Rolle zukommt. Diese Hypothese wäre allerdings in weiterführenden Projekten zu untersuchen, und geeignete Instrumente zur Klärung dieses Zusammenhangs wären wichtig. Das Instrument leistet einen Beitrag für geplante Forschungsprojekte, in denen erwartete Zusammenhänge von Professionswissen hinsichtlich verschiedener Zielvariablen untersucht werden können.

Die Ergebnisse, die mit dem Befragungsinstrument gewonnen wurden, sind nicht zuletzt für die betroffenen Kinder mit Leseschwierigkeiten praktisch relevant. Falls sich fachliche und fachdidaktische Wissenslücken bei Lehrpersonen feststellen ließen und Leistungsunterschiede von Schülerinnen und Schülern damit erklärt werden könnten, wären u.a. Ausbildungsinstitutionen dazu aufgefordert, diesen Defiziten in Zukunft entgegenzuwirken. Der Einsatz des Instruments lässt damit Schlüsse darüber zu, wo in der Ausbildung zum Thema Anfangsleseunterricht besondere Anstrengungen notwendig sind. Das Projekt leistet damit einen Beitrag „um die für ein optimales Lernen der Schülerinnen und Schüler notwendige Kompetenzen der Lehrpersonen theoretisch zu beschreiben, empirisch zu erfassen und in die Richtung eines gewünschten Zustands zu lenken“ (Maag Merki & Werner, 2014, S. 745).

Ein bedeutender Beitrag des vorliegenden Projekts besteht darin, dass zentrale Wissensbereiche operationalisiert und messbar gemacht werden, über die angehende und ausgebildete Lehrpersonen und SHP Bescheid wissen sollten. Es werden fachliche und fachdidaktische Inhalte definiert, über die Lehrpersonen verfügen müssen, um guten Anfangsleseunterricht zu gestalten und um Hürden, die sich bei Kindern in dieser Phase stellen, zuverlässig zu erkennen (Kap. 2). Von einer Lehrperson am Ende der Ausbildung und mit einigen Jahren Berufserfahrung im entsprechenden Fach kann und muss dies erwartet werden. Vor allem aber müssen Expertinnen und Experten wie etwa ausgebildete Schulische Heilpädagoginnen und Heilpädagogen mit Berufserfahrung dazu in der Lage sein. Das vorliegende Instrument zeigt, dass sich dieses Wissen – zumindest bei der Stichprobe der SHP – reliabel und valide messen lässt und dass sich das Wissen von Lehrpersonen ohne Ausbildung in Schulischer Heilpädagogik im Vergleich dazu quantitativ unterscheidet.

Des Weiteren wird mit dem vorliegenden Projekt die Forderung eingelöst, ein Instrument zu entwickeln, das an empirisch bewährten Modellen zum Professionswissen anschließt und damit einen „Diskurs über Befunde auf breiter Ebene“ ermöglicht (Baumert, 2017, S. 7). Dies gilt insofern, als sich die theoretische Konzeption und Operationalisierung der Wissensdimensionen stark an bestehenden Instrumenten anderer Domänen orientiert (Blömeke et al., 2013a;

Krauss et al., 2017a; Kunter et al., 2011a). Bei der hier präsentierten Studie handelt es sich um eine Art „Grundlagenforschung“ mit dem Anliegen, mehr über die Art und Struktur des untersuchten Wissens zu erfahren (Lindl & Krauss, 2017; Wiprächtiger-Geppert et al., 2015).

Die Anwendung des Instruments kann – im Sinne eines nützlichen Bausteins – helfen, spezifische Fragen nach der Ausgestaltung der Lehrerinnen- und Lehrerbildung zu klären. Der Beitrag könnte beispielsweise darin bestehen, den aktuellen Wissensstand auf Gruppenebene objektiv zu erfassen und für vergleichende Analysen verschiedener Ausbildungsinstitutionen heranzuziehen und auf das jeweilige Curriculum zu beziehen (Einschränkungen siehe weiter oben). Denkbar wäre auch der Einsatz für den längsschnittlichen Vergleich von Leistungszuwächsen über verschiedene Ausbildungsphasen hinweg. „Durch Kompetenzmessungen über die Phasen der Lehrerbildung hinweg kann sichtbar gemacht werden, was gelernt wird und wie dies mit gleichzeitigen und nachfolgenden Erfolgskriterien einhergeht“ (Spinath, 2019, S. 18).

Mit der Entwicklung des Instruments ist eine Grundlage geschaffen, die zur Evaluation von Aus- und Weiterbildungsgängen dienen kann. Vor der überhöhten Erwartung, die Wirksamkeit einer Ausbildung umfänglich messen zu können, muss angesichts der Komplexität und der Vielzahl der beteiligten Faktoren aber gewarnt werden (Terhart, 2012). Ein Test kann die unterrichtliche Praxis nie abbilden, er kann lediglich ein Indikator für die Kompetenz sein (Schlicher & Krauss, 2019). Damit ist zumindest ein wichtiger Indikator – nämlich das Wissen – genauer beschrieben und operationalisiert.

Der Nutzen der vorliegenden Ergebnisse liegt also auch darin, die Weiterentwicklung der Lehrerbildung voranzutreiben. Der Blick kann dabei auf die Ebene der Institutionen, aber auch direkt auf Beurteilungs- und Selektionsprozesse der Studierenden gelenkt werden. Richtet man den Fokus auf letzteres, liegt der Nutzen darin, auf eine objektivere Entscheidungsgrundlage für die Beurteilung von Studierenden zurückgreifen zu können. Dieses Postulat wurde etwa im Rahmen der Entwicklung von Standards in der Lehrerinnen- und Lehrerbildung schon vor Jahren angestrebt (Oser, 1997). Eine noch zu bearbeitende Frage betrifft die Bedeutung der Selektion und Beurteilung von Studierenden für die Qualitätssteigerung der Ausbildung von Lehrpersonen (Spinath, 2019, 24). Erhebungsverfahren, die eine valide Einschätzung und Beurteilung ermöglichen, sind in diesem Auswahlprozess von hoher Relevanz (Maag Merki & Werner, 2014). Auch hier leistet die vorliegende Arbeit einen Beitrag im Sinne einer objektiven vergleichbaren Beurteilung von Wissen und Können zu einem eingegrenzten und wichtigen Wissensbereich.

9 Literaturverzeichnis

- Adams, R., Wu, M. & Wilson, M. R. (2015). *ACER ConQuest: Generalised Item response Modelling Software (Version 4.5.2)*. University of California, Berkeley.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5. Aufl.). Arlington: American Psychiatric Association.
- Annevirta, T., Laakkonen, E., Kinnunen, R. & Vauras, M. (2007). Developmental dynamics of metacognitive knowledge and text comprehension skill in the first primary school years. *Metacognition Learning*, 2, 21-39.
- Antonietti, J.-P., Moreau, J., Moser, U. & Ramseier, E. (2006). *Harmos. Zum methodologischen Vorgehen*. [Wegleitung der Methodengruppe. Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK)]. Verfügbar unter: www.edk.ch [Zugriff: Dezember 2015].
- Artelt, C. (2009). Diagnostische Urteile von Lehrkräften im Bereich der Lesekompetenz. In A. Bertschi-Kaufmann & C. Rosebrock (Hrsg.), *Literalität. Bildungsaufgabe und Forschungsfeld* (S. 125-136). Weinheim: Juventa.
- Artelt, C., Naumann, J. & Schneider, W. (2010). Lesemotivation und Lernstrategien. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009 Bilanz nach einem Jahrzehnt* (S. 73-112). Münster: Waxmann.
- Asadi, I., A. & Ibrahim, R. (2018). The simple view of reading Model in the transparent and deep versions of arabic orthographie. *Reading Psychologie*, 39(6), 537-552.
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417-422.
- Baer, M., Dörr, G., Fraefel, U., Kocher, M., Küster, O., Larcher, S. et al. (2007). Werden angehende Lehrpersonen durch das Studium kompetenter? Kompetenzaufbau und Standarderreichung in der berufswissenschaftlichen Ausbildung an drei Pädagogischen Hochschulen in der Schweiz und in Deutschland. *Unterrichtswissenschaft*, 35(1), 15-47.
- Baer, M., Kocher, M., Wyss, C., Guldemann, T., Larcher, S. & Dörr, G. (2011). Lehrerbildung und Praxiserfahrung im ersten Berufsjahr und ihre Wirkung auf die Unterrichtskompetenzen von Studierenden und jungen Lehrpersonen im Berufseinstieg. *Zeitschrift für Erziehungswissenschaft*, 14(1), 85-117.
- Ball, D. L., Thames, M. H. & Phelps, G. (2008). Content knowledge for teaching: what makes it special? *Journal of Teacher Education*, 59(5), 389-407.
- Bamberger, R. & Vanecek, E. (1984). *Lesen – Verstehen – Lernen – Schreiben. Die Schwierigkeitsstufen von Texten in deutscher Sprache*. Wien, Frankfurt am Main: Jugend und Volk.
- Basilicata, A. & Meier, A. (2011). *Leseförderung an der Mittelstufe*. Unveröffentlichtes Manuskript, Interkantonale Hochschule für Heilpädagogik Zürich.
- Baumert, J. (2017). Vorwort. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. Mulder (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik* (S. 7). Münster: Waxmann.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469-520.
- Baumert, J. & Kunter, M. (2011a). Das Kompetenzmodell von COACTIV. In J. Baumert, M. Runter, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29-53). Münster: Waxmann.
- Baumert, J. & Kunter, M. (2011b). Das mathematikspezifische Wissen von Lehrkräften, kognitive Aktivierung im Unterricht und Lernfortschritte von Schülerinnen und Schülern. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 163-192). Münster: Waxmann.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A. et al. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress *American Educational Research Journal*, 47(1), 133-180.

- Behrmann, L. & Souvignier, E. (2013). The relation between teachers' diagnostic sensitivity, their instructional activities, and their students' achievement gains in reading. *Zeitschrift für pädagogische Psychologie*, 27(4), 283-293.
- Bellenberg, G. & Korte, J. (2019). Lehrerbildung gestalten: Der Beitrag erziehungswissenschaftlicher Theorien und Forschungsansätze zur Professionalisierung angehender Lehrerinnen und Lehrer. In N. McElvany, W. Bos, H. G. Holtappels & A. Ohle-Peters (Hrsg.), *Bedingungen und Effekte von Lehrerbildung, Lehrkraftkompetenzen und Lehrkrafthandeln. Dortmunder Symposium der Empirischen Bildungsforschung* (S. 51-65). Münster: Waxmann.
- Benkmann, R. (2005). Zur Veränderung sonderpädagogischer Professionalität im Gemeinsamen Unterricht aus Sicht der Pädagogik bei Lernbeeinträchtigungen. *Zeitschrift für Heilpädagogik*, 56(10), 418 - 427.
- Berendes, K., Schnitzler, C. D., Willmes, K. & Huber, W. (2010). Die Bedeutung von Phonembewusstheit und semantisch-lexikalischen Fähigkeiten für Schriftsprachleistungen in der Grundschule 1: The Importance of Phonemic Awareness and Vocabulary for Written Language Performance in Elementary School. *Sprache Stimme Gehör*, 34(3), 48-55.
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35(5), 463-482.
- Bertschi-Kaufmann, A. (2007). Computer oder Buch? – Beides! Leseförderung im Spannungsfeld zwischen traditionellen Textarten und multimedialen Formaten. In H. Mitzlaff (Hrsg.), *Internationales Handbuch Computer (ICT), Grundschule, Kindergarten und Neue Lernkultur* (Bd. 529-535). Hohengehren: Schneider Verlag.
- Bertschi-Kaufmann, A. (2011). Leseverhalten beobachten – Lesen und Schreiben in der Verbindung. In A. Bertschi-Kaufmann (Hrsg.), *Lesekompetenz Leseleistung Leseförderung. Grundlagen, Modelle und Materialien* (4. Aufl., S. 96-108). Zug: Klett, Kallmeyer.
- Bertschi-Kaufmann, A., Hagendorf, P., Kruse, G., Rank, K., Riss, M. & Sommer, T. (2007). *Lesen. Das Training. Lesefertigkeiten – Lesegeläufigkeit – Lesestrategien. Stufe I und II*. Seelze: Friedrich.
- Bertschi-Kaufmann, A. & Kappeler, S. (2010). Gegenwärtiger Stand der empirischen Unterrichtsforschung zur Vermittlung von Lesekompetenz. In M. Kämper- van den Boogaart & K. H. Spinner (Hrsg.), *Deutschunterricht in Theorie und Praxis. (DTP). Lese- und Literaturunterricht (Teil 2). Kompetenzen und Unterrichtsziele. Methoden und Unterrichtsmaterialien. Gegenwärtiger Stand der empirischen Unterrichtsforschung* (S. 275-305). Baltmannsweiler: Schneider-Verlag Hohengehren.
- Bertschi-Kaufmann, A. & Schneider, H. (2006). Entwicklung von Lesefähigkeit: Massnahmen – Messungen – Effekte. Ergebnisse und Konsequenzen aus dem Forschungsprojekt "Lese- und Schreibkompetenzen fördern". *Schweizerische Zeitschrift für Bildungswissenschaften*, 28(3), 393-424.
- Binks-Cantrell, E., Joshi, M. & Washburn, E. K. (2012). Validation of an instrument for assessing teacher knowledge of basic language constructs of literacy. *Annals of Dyslexia*, 62(3), 153-171.
- Blömeke, S. (2011). Teacher Education and Development Study: Learning to Teach (TEDS-LT) - Erfassung von Lehrerkompetenzen in gering strukturierten Domänen. In S. Blömeke, A. Bremerich-Vos, H. Haudeck, G. Kaiser, G. Nold, K. Schwippert & H. Willenberg (Hrsg.), *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT* (S. 7-24). Münster: Waxmann.
- Blömeke, S. (2013). Einleitung: Professionelle Kompetenz im Studienverlauf. In S. Blömeke, A. Bremerich-Vos, G. Kaiser, G. Nold, H. Haudeck, J.-U. Kessler & K. Schwippert (Hrsg.), *Professionelle Kompetenzen im Studienverlauf. Weitere Ergebnisse zur Deutsch-, Englisch- und Mathematiklehrausbildung aus TEDS-LT* (S. 7-24). Münster: Waxmann.
- Blömeke, S., Bremerich-Vos, A., Haudeck, H., Kaiser, G., Nold, G., Schwippert, K. et al. (2011). *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT*. Münster: Waxmann.
- Blömeke, S., Bremerich-Vos, A., Kaiser, G., Nold, G., Haudeck, H., Kessler, J.-U. et al. (2013a). *Professionelle Kompetenzen im Studienverlauf. Weitere Ergebnisse zur Deutsch-, Englisch- und Mathematiklehrausbildung aus TEDS-LT*. Münster: Waxmann.

- Blömeke, S., Buchholtz, C. & Bremerich-Vos, A. (2013b). Zusammenhang institutioneller Merkmale im dem Wissenserwerb im Lehramtsstudium. In S. Blömeke, A. Bremerich-Vos, G. Kaiser, G. Nold, H. Haudeck, J.-U. Kessler & K. Schwippert (Hrsg.), *Professionelle Kompetenzen im Studienverlauf. Weitere Ergebnisse zur Deutsch-, Englisch- und Mathematiklehrausbildung aus TEDS-LT* (S. 167-187). Münster: Waxmann.
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015a). Beyond dichotomies. Competence viewed as a continuum. *Zeitschrift für Psychologie*, 223(1), 3-13.
- Blömeke, S., Kaiser, G. & Lehmann, R. (2008). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare. Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung*. Münster: Waxmann.
- Blömeke, S., Kaiser, G. & Lehmann, R. (2010a). *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann.
- Blömeke, S., Kaiser, G. & Lehmann, R. (2010b). *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich*. Münster: Waxmann.
- Blömeke, S. & König, J. (2010). Messung des pädagogischen Wissens: Theoretischer Rahmen und Teststruktur. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich* (S. 239-263). Münster: Waxmann.
- Blömeke, S., König, J., Suhl, U., Hoth, J. & Döhrmann, M. (2015b). Wie situationsbezogen ist die Kompetenz von Lehrkräften? Zur Generalisierbarkeit der Ergebnisse von videobasierten Performanztests. *Zeitschrift für Pädagogik*, 61(3), 310-327.
- Blömeke, S., Seeber, S., Kaiser, G., Schwarz, B., Lehmann, R., Felbrich, A. et al. (2009). Differentielle Item-Analyse zur Entwicklung professioneller Kompetenz angehender Lehrkräfte während der Lehrerausbildung. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (S. 311-327). Weinheim: Beltz.
- Bochnik, K. (2017). *Sprachbezogene Merkmale als Erklärung für Disparitäten mathematischer Leistung. Differenzierte Analysen im Rahmen einer Längsschnittstudie in der dritten Jahrgangsstufe*. Münster: Waxmann.
- Böhm-Kasper, O. & Weishaupt, H. (2008). Quantitative Ansätze und Methoden in der Schulforschung. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (2, S. 91-123). Wiesbaden: Springer VS.
- Böhme, K., Bremerich-Vos, A. & Robitzsch, A. (2009). Aspekte der Kodierung von Schreibaufgaben. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 290-329). Weinheim: Beltz.
- Bond, T. G. & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah: Lawrence Erlbaum.
- Borsboom, D. (2006). When Does Measurement Invariance Matter? *Medical Care*, 44(11), 176-181.
- Bortz, J. & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (3. überarbeitete Aufl.). Heidelberg: Springer.
- Bortz, J. & Lienert, G. A. (2003). *Kurzgefasste Statistik für klinische Forschung. Leitfaden für die verteilungsfreie Analyse kleiner Stichproben* (2. aktualisierte und bearbeitete Aufl.). Berlin: Springer.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7., vollständig überarbeitete und erweiterte Aufl.). Berlin: Springer.
- Bräuer, C. R. (2010). *Könnerschaft und Kompetenz in der Leseausbildung. Theoretische und empirische Perspektiven*. Weinheim: Juventa.
- Bremerich-Vos, A., Buchholtz, C. & König, J. (2016). *PlanvoLL-D – Die Bedeutung des professionellen Wissens angehender Deutschlehrkräfte für ihre Planung von Unterricht: Validierung und methodische Innovation (KoKoHs Working Paper 10)*. Verfügbar unter: https://www.kompetenzen-im-hochschulsektor.de/files/2018/05/WP10_final_032016.pdf [Zugriff: Juni 2020].

- Bremerich-Vos, A. & Dämmer, J. (2013). Professionelles Wissen im Studienverlauf: Lehramt Deutsch. In S. Blömeke, A. Bremerich-Vos, G. Kaiser, G. Nold, H. Haudeck, J.-U. Kessler & K. Schwippert (Hrsg.), *Professionelle Kompetenzen im Studienverlauf. Weitere Ergebnisse zur Deutsch-, Englisch- und Mathematiklehrerbildung aus TEDS-LT* (S. 47-75). Münster: Waxmann.
- Bremerich-Vos, A., Dämmer, J., Willenberg, H. & Schwippert, K. (2011). Professionelles Wissen von Studierenden des Lehramts Deutsch. In S. Blömeke, A. Bremerich-Vos, H. Haudeck, G. Kaiser, G. Nold, K. Schwippert & H. Willenberg (Hrsg.), *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT* (S. 47-99). Münster: Waxmann.
- Bromme, R. (1992). *Der Lehrer als Experte. Zur Psychologie des professionellen Wissens*. Bern: Huber.
- Bruckmaier, G., Krauss, S., Leiss, D., Blum, W., Neubrand, M. & Brunner, M. (2013). COACTIV-Video. Eine unterrichtsnahe Erfassung fachdidaktischen Wissens mittels Videovignetten. In *Beiträge zum Mathematikunterricht. Band 1* (S. 212-215). Münster: WTM.
- Brügelmann, H. (2014). Kommentare zu Corvacho del Toro & Thomé (2013): Zum Effekt des Fachwissens von Lehrkräften auf die Rechtschreibleistung von Grundschulern. *Lernen und Lernstörungen*, 3(1), 59-66.
- Brühwiler, C., Hollenstein, L., Affolter, B., Biedermann, H. & Oser, F. (2017). Welches Wissen ist unterrichtsrelevant? Prädiktive Validität dreier Messinstrumente zur Erfassung des pädagogisch-psychologischen Wissens von Lehrpersonen. *Zeitschrift für Bildungsforschung*, 7(3), 209-228.
- Brühwiler, C., Ramseier, E. & Steinmann, S. (2015). Vorbildung oder Ausbildung? Zum Erwerb mathematischen und mathematikdidaktischen Wissens in der Lehrpersonenausbildung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 33(1), 22-45.
- Brunner, M., Kunter, M., Krauss, S., Baumert, J., Blum, W., Dubberke, T. et al. (2006). Welche Zusammenhänge bestehen zwischen dem fachspezifischen Professionswissen von Mathematiklehrkräften und ihrer Ausbildung sowie beruflichen Fortbildung? *Zeitschrift für Erziehungswissenschaft*, 9(4), 521-544.
- Buchholtz, N. & Kaiser, G. (2013). Professionelles Wissen im Studienverlauf: Lehramt Mathematik. In S. Blömeke, A. Bremerich-Vos, G. Kaiser, G. Nold, H. Haudeck, J.-U. Kessler & K. Schwippert (Hrsg.), *Professionelle Kompetenzen im Studienverlauf. Weitere Ergebnisse zur Deutsch-, Englisch- und Mathematiklehrerbildung aus TEDS-LT*. (S. 107-143). Münster: Waxmann.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. aktualisierte Aufl.). München: Pearson.
- Bühner, M. & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson Studium.
- Bundesverband Legasthenie und Dyskalkulie. (2018). *Legasthenie. Ratgeber zum Thema Legasthenie - Erkennen und Verstehen* [Informationsschrift des Bundesverband Legasthenie und Dyskalkulie (12. Aufl.)]. Verfügbar unter: www.bvl-legasthenie.de [Zugriff: Mai 2020].
- Cain, K. & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal*, 11, 489-503.
- Carlisle, J. F., Correnti, R., Phelps, G. & Zeng, J. (2009). Exploration of the contribution of teachers' knowledge about reading to their students' improvement in reading. *Reading and Writing*, 22(4), 447-486.
- Carlisle, J. F., Kelcey, B., Rowan, B. & Phelps, G. (2011). Teachers' Knowledge About Early Reading: Effects on Students' Gains in Reading Achievement. *Journal of Research on Educational Effectiveness* 4(4), 289-321.
- Christ, O. & Schlüter, E. (2012). *Strukturgleichungsmodelle mit Mplus: eine praktische Einführung*. München: Oldenbourg Verlag.
- Christmann, U. & Groeben, N. (1999). Psychologie des Lesens. In B. Franzmann, K. Hasemann, D. Löffler & E. Schön (Hrsg.), *Handbuch Lesen* (S. 145-223). München: Saur.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2. Ed.). Hillsdale: Lawrence Erlbaum.

- Cole, D. A., Perkins, C. E. & Zelkowitz, R. L. (2016). Impact of homogeneous and heterogeneous parceling strategies when latent variables represent multidimensional constructs. *Psychological Methods*, 21(2), 164-174.
- Coltheart, M. (2005). Modeling Reading: The Dual-Route Approach. In M. J. Snowling & C. Hulme (Hrsg.), *The Science of Reading: A Handbook* (S. 6-23). Malden: Blackwell Publishing Ltd.
- Combe, A. & Kolbe, F.-U. (2008). Lehrerprofessionalität: Wissen, Können, Handeln. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (S. 857-875). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Corvacho del Toro, I. M. (2013). *Fachwissen von Grundschullehrkräften. Effekt auf die Rechtschreibleistung von Grundschulern* [Dissertation Universität Bamberg]. Verfügbar unter: <https://d-nb.info/1058948733/34> [Zugriff: Juni 2020].
- Corvacho del Toro, I. M. & Thomé, G. (2013). Zum Effekt des Fachwissens von Lehrkräften auf die Rechtschreibleistung von Grundschulern. *Lernen und Lernstörungen*, 2(1), 21-33.
- Crämer, C., Füssenich, I. & Schumann, G. (1996). Lese- und Schreibschwierigkeiten im Zusammenhang mit Problemen der gesprochenen Sprache. *Die Sprachheilarbeit*, 41(1), 5-21.
- Depaepe, F., Verschaffel, L. & Klechtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, 34, 12-25.
- Deutsche Gesellschaft für Kinder- und Jugendpsychiatrie Psychosomatik und Psychotherapie. (2015). *Diagnostik und Behandlung von Kindern und Jugendlichen mit Lese- und / oder Rechtschreibstörung. Evidenz- und konsensbasierte Leitlinie. AWMF-Registernummer 028 - 044*. Verfügbar unter: https://www.bvl-legalasthenie.de/images/static/pdfs/Leitlinien/LF_Leitlinie.pdf [Zugriff: Mai 2020].
- Deutsches Institut für Medizinische Dokumentation und Information (DIMDI). (2019). *Wann kommt die ICD-11?* Verfügbar unter: <https://www.dimdi.de/dynamic/de/faq/faq/Wann-kommt-die-ICD-11/> [Zugriff: November 2019].
- Diehl, K. & Hartke, B. (2012). *IEL-1. Inventar zur Erfassung der Lesekompetenz im 1. Schuljahr. Ein curriculumbasiertes Verfahren zur Abbildung des Lernfortschritts*. Göttingen: Hogrefe.
- Dilling, H., Mombour, W., Schmidt, M. & Schulte-Markwort, E. (2011). *Internationale Klassifikation psychischer Störungen. ICD-10 Kapitel V (F). Diagnostische Kriterien für Forschung und Praxis* (5. Aufl.). Bern: Huber.
- Döhrmann, M., Kaiser, G. & Blömeke, S. (2010). Messung des mathematischen und mathematikdidaktischen Wissens: Theoretischer Rahmen und Teststruktur. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich* (S. 169-194). Münster: Waxmann.
- Dörfler, T., Schmitt, M. & Rutsch, J. (2015). *EKoL 3 - Lesekompetenz: Erfassung von fachdidaktischem Wissen in Bezug auf Lesekompetenz bei angehenden Lehrkräften* Verfügbar unter: <https://www.ph-heidelberg.de/fun-ekol/startseite/projekt/projektphase-1-2013-2016/ekol-3-lesekompetenz.html> [Zugriff: Juni 2020].
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. vollständig überarbeitete, aktualisierte Aufl.). Berlin: Springer.
- Downing, S., M. (2006). Twelve Steps for Effective Test Development. In S. Downing, M. & T. Haladyna, M. (Hrsg.), *Handbook of Test Development* (S. 3-25). Mahawah: Routledge.
- Draba, R. E. (1977). *The identification and interpretation of item bias. Educational Statistics Laboratory - Memo 25*. Verfügbar unter: www.rasch.org/memo25.htm [Zugriff: Juni 2020].
- Du Yates, F. (1995). *When to adjust for Differential Item Functioning. Rasch Measurement Transactions*. Verfügbar unter: <https://www.rasch.org/rmt/rmt91e.htm> [Zugriff: Juni 2020].
- Ehri, L. C. (1997). Learning to read and learning to spell are one and the same, almost. In C. A. Perfetti, L. Rieben & M. Fayol (Hrsg.), *Learning to spell. Research, theory and practice across languages* (S. 237-269). Mahwah: Lawrence Erlbaum.
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z. & Shanahan, T. (2001). Systematic Phonics Instruction Helps Students Learn to Read: Evidence from the National Reading Panel's Meta-Analysis. *Reading Research Quarterly*, 36(3), 250-287.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2015). *Statistik und Forschungsmethoden* (4. korrigierte Aufl.). Weinheim: Beltz.

- Eid, M. & Schmidt, K. (2014). *Testtheorie und Testkonstruktion*. Göttingen: Hogrefe.
- Embretson, S. E. & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah: Lawrence Erlbaum.
- Ennemoser, M., Marx, P. & Schneider, W. (2012). Spezifische Vorläuferfertigkeiten der Lesegeschwindigkeit, des Leseverständnisses und des Rechtschreibens: Evidenz aus zwei Längsschnittstudien vom Kindergarten bis zur 4. Klasse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 44(2), 53-67.
- Esser, G. & Schmidt, M. (1993). Die langfristige Entwicklung von Kindern mit Lese-Rechtschreibschwäche. *Zeitschrift für Klinische Psychologie*, 22(2), 100-116.
- Falkai, P. & Wittchen, H.-U. (2015). *Diagnostische Kriterien DSM-5® (American Psychiatric Association)*. Göttingen: Hogrefe.
- Ferber, N. (2014). *Kompetenzentwicklung im Fachwissen Chemie in der Sekundarstufe I*. Kiel: Logos Verlag.
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4. Ed.). Los Angeles: SAGE.
- Fischer, U. & Gasteiger-Klicpera, B. (2013). Prävention von Leseschwierigkeiten. Diagnose und Förderung im Anfangsunterricht. *Didaktik Deutsch. Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur*, 19(35), 63-81.
- Forster, M. & Martschinke, S. (2003). *Leichter lesen und schreiben lernen mit der Hexe Susi*. Donauwörth: Auer.
- Frey, A. (2006). Methoden und Instrumente zur Diagnose beruflicher Kompetenzen von Lehrkräften. Eine erste Standortbestimmung zu bereits publizierten Instrumenten. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern Ausbildung und Beruf* (S. 30-46). Weinheim: Beltz.
- Frith, U. (1985). Beneath the Surface of Developmental Dyslexia. In K. Patterson, J. C. Marshall & M. Coltheart (Hrsg.), *Surface dyslexia: neuropsychological and cognitive studies of phonological reading* (S. 301-330). London: Lawrence Erlbaum.
- Füssenich, I. & Löffler, C. (2008). *Schriftspracherwerb. Einschulung, erstes und zweites Schuljahr* (2. durchgesehene Aufl.). München: Reinhardt.
- Gäde, C. J., Schermelleh-Engel, K. & Brandt, H. (2020). Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. vollständig neu bearbeitete, erweiterte und aktualisierte Aufl., S. 615-659). Berlin: Springer.
- Galuschka, K. & Schulte-Körne, G. (2015). Evidenzbasierte Interventionsansätze und forschungsbasierte Programme zur Förderung der Leseleistung bei Kindern und Jugendlichen mit Lesestörung – Ein systematischer Review. *Zeitschrift für Erziehungswissenschaft*, 18, 473-487.
- Galuschka, K. & Schulte-Körne, G. (2016). Diagnostik und Förderung von Kindern und Jugendlichen mit Lese- und/oder Rechtschreibstörung. Klinische Leitlinien. *Deutsches Ärzteblatt*, 113(16), 279-286.
- Geiser, C. (2010). *Datenanalyse mit Mplus. Eine anwendungsorientierte Einführung*. Wiesbaden: VS Verlag, Springer.
- Geiser, C. & Eid, M. (2010). Item-Response-Theorie. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse*. (S. 311–332). Wiesbaden: VS Verlag für Sozialwissenschaften
- Gold, A. (2009). Leseflüssigkeit. Dimensionen und Bedingungen bei leseschwachen Hauptschülern. In A. Bertschi-Kaufmann & C. Rosebrock (Hrsg.), *Literalität. Bildungsaufgabe und Forschungsfeld* (S. 151-164). Weinheim: Juventa.
- Gold, A., Mokhlesgerami, J., Rühl, K., Schreblowski, S. & Souvignier, E. (2006). *Wir werden Textdetektive – Lehrermanual und Arbeitsheft* (2. Aufl.). Göttingen: Vandenhoeck & Ruprecht.
- Gold, A., Nix, D., Rieckmann, C. & Rosebrock, C. (2010). Bedingungen des Textverstehens bei leseschwachen Zwölfjährigen mit und ohne Zuwanderungshintergrund. *Didaktik Deutsch. Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur*, 16(28), 59-74.
- Gold, A., Trenk-Hinterberger, I. & Souvignier, E. (2009). „Die Textdetektive“ – Ein strategieorientiertes Programm zur Förderung des Leseverständnisses. In W. Lenhard & W. Schneider (Hrsg.), *Diagnostik und Förderung des Leseverständnisses*.

- Goldhammer, F. & Hartig, J. (2012). Interpretation von Testresultaten und Testeichung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. durchgesehen und aktualisierte Aufl., S. 173-201). Heidelberg: Springer.
- Gorecki, B. & Landerl, K. (2015). Phonologische Bewusstheit. Ist die phonologische Bewusstheit ein Prädiktor für die Leseleistung? *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie*, 47(3), 139-146.
- Götze, B., Hasselhorn, M. & Kiese-Himmel, C. (2000). Phonologisches Arbeitsgedächtnis, Wortschatz und morpho-syntaktische Sprachleistungen im Vorschulalter. *Sprache und Kognition*, 19(1/2), 15-21.
- Gough, P. B. & Tunmer, W. E. (1986). Decoding, Reading and Reading Disability. *Remedial and Special Education*, 7(1), 6-10.
- Grand, A., Dittrich, R. & Hatzinger, R. (2017). *Präferenzmodell in der Praxis. Analyse von Paarvergleichen, Likert Items und Ranking mit R-prefmod*. Wien: Facultas.
- Groeben, N. & Christmann, U. (1996). Textverstehen und Textverständlichkeit aus sprach-/denkpsychologischer Sicht. In W. Börner & K. Vogel (Hrsg.), *Texte im Fremdspracherwerb: Verstehen und Produzieren* (S. 67-89). Tübingen: Narr.
- Grollmann, P. & Jude, N. (2008). Kompetenz in der empirischen Bildungsforschung und in der Beruflichen Bildung - Anknüpfungspunkte für einen hoffentlich fruchtbaren Dialog. In N. Jude, J. Hartig & E. Klieme (Hrsg.), *Kompetenzerfassung in pädagogischen Handlungsfeldern. Theorien, Konzepte und Methoden* (S. 141-148). Bonn: Bundesministerium für Bildung und Forschung.
- Günther, H. (1986). Ein Stufenmodell der Entwicklung kindlicher Lese- und Rechtschreibstrategien. In H. Brügelmann (Hrsg.), *ABC und Schriftsprache: Rätsel für Kinder, Lehrer und Forscher* (S. 32-54). Konstanz: Faude.
- Günther, H. (2007). *Schriftspracherwerb und LRS. Methoden, Förderdiagnostik und praktische Hilfen*. Weinheim: Beltz.
- Guthrie, J. T. & Wigfield, A. (2000). Engagement and Motivation in Reading. In M. L. Kamil, P. B. Mosenthal, D. P. Pearson & R. Barr (Hrsg.), *Handbook of Reading Research* (S. 403-422). Mahwah: Lawrence Erlbaum.
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park: Sage Publications.
- Hanke, P., König, J., Jäger-Biela, D., Pohl, T., Schabmann, A., Becker-Mrotzek, M. et al. (2019). Professionelles Wissen von Lehramtsstudierenden zum basalen Lesen- und Schreibenlernen – ein interdisziplinäres Projekt. In C. Donie, F. Foerster, M. Obermayr, A. Deckwerth, G. Kammermeyer, G. Lenske, M. Leuchter & A. Wildemann (Hrsg.), *Grundschulpädagogik zwischen Wissenschaft und Transfer*. Wiesbaden: Springer VS.
- Hartig, J. (2009). Messung der Kompetenzen von Lehrpersonen mit Modellen der Item-Response-Theorie. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (S. 295-310). Weinheim: Beltz.
- Hartmann, E. (2002). *Möglichkeiten und Grenzen einer präventiven Intervention zur phonologischen Bewusstheit von lautsprachgestörten Kindergartenkindern*. Fribourg: Sprachimpuls.
- Hartmann, E. (2006). *In Bildern denken – Texte besser verstehen. Lesekompetenz strategisch stärken*. München: Reinhardt.
- Hartmann, E. (2007). Erfolg versprechende Prävention von Leseschwierigkeiten in Kindergarten und Primarschule: Ein Überblick. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, 76(2), 114-127.
- Hartmann, E. (2013). Schulweite Prävention von Lese-Rechtschreibschwierigkeiten im RTI-Modell. Ein Überblick. *Gemeinsam leben*(2), 100-108.
- Hartmann, E. & Niedermann, A. (2008). Förderung der Leseflüssigkeit. *Schweizerische Zeitschrift für Heilpädagogik*, 14(4), 44-47.
- Hattie, J. (2003). *Teachers Make a Difference. What is the research evidence? Paper presented at the Building Teacher Quality: What does the research tell us (ACER Research Conference, Melbourne, Australia)*. Verfügbar unter: https://research.acer.edu.au/research_conference_2003/4/ [Zugriff: Juni 2020].

- Hayes, A. F. & Krippendorff, K. (2007). Answering the Call for a Standard. Reliability Measure for Coding Data. *Communication Methods and Measures*, 1(1), 77-89.
- Hefti, C. & Ventocilla Metz, L. (2009). *Zum Wissensstand von 1./2. Klasse-Lehrpersonen über Schriftspracherwerb und Lese-Rechtschreibschwierigkeiten: Eine Befragung in der deutschsprachigen Schweiz*. Unveröffentlichte Bachelorarbeit. Universität Freiburg/CH.
- Heine, J.-H. & Tarnai, C. (2015). Pairwise Rasch model item parameter recovery under sparse data conditions. *Psychological Test and Assessment Modeling*, 57(1), 3-36.
- Helbig, P., Kirschhock, E.-M., Martschinke, S. & Kummer, U. (2005). *Schriftspracherwerb im entwicklungsorientierten Unterricht. Lernwege bereiten und begleiten*. Bad Heilbrunn: Klinkhardt.
- Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (4. aktualisierte Aufl.). Seelze-Velber: Klett Kallmeyer.
- Helmke, A. (2017). *Unterrichtsqualität und Lehrerprofessionalität : Diagnose, Evaluation und Verbesserung des Unterrichts* (7. überarbeitete Auflage). Seelze-Velber: Klett, Kallmeyer.
- Hepberger, B., Lindmeier, A., Moser Opitz, E. & Heinze, A. (2017). „Zähl’ nochmal genauer!“ – Handlungsnahe mathematikbezogene Kompetenzen von pädagogischen Fachkräften erheben. In S. Schuler, C. Streit & G. Wittmann (Hrsg.), *Perspektiven mathematischer Bildung im Übergang vom Kindergarten zur Grundschule* (S. 239-253). Wiesbaden: Springer.
- Herné, K.-L. & Löffler, C. (2014). *LRS: Schwierigkeiten erkennen – Fähigkeiten fördern. Ein Praxishandbuch für Lehrende der Klassen 1-6*. Seelze: Kallmeyer und Klett.
- Herzmann, P. (2010). Die Förderung der Lesekompetenz in Hauptschulbildungsgängen. Eine komparative Analyse von Unterrichtseinheiten und Unterrichtsgesprächen zum Leseverstehen und zur Lesemotivation. In J. König & B. Hofmann (Hrsg.), *Professionalität von Lehrkräften. Was sollen Lehrkräfte im Lese- und Schreibunterricht wissen und können?* (S. 270-281). Berlin: Deutsche Gesellschaft für Lesen und Schreiben.
- Hildebrandt, A., Jäckle, S., Wolf, F. & Heindl, A. (2015). *Methodologie, Methoden, Forschungsdesign: Ein Lehrbuch für fortgeschrittene Studierende der Politikwissenschaft*. Wiesbaden: Springer VS.
- Hill, H. C., Ball Loewenberg, D. & Schilling, G. S. (2008). Unpacking Pedagogical Content Knowledge: Conceptualizing and Measuring Teachers' Topic-Specific Knowledge of Students. *Journal for Research in Mathematics Education*, 39(4), 372-400.
- Hill, H. C., Schilling, S. G. & Loewenberg Ball, D. (2004). Developing Measures of Teachers' Mathematics Knowledge for Teaching. *The Elementary School Journal*, 105(1), 11-33.
- Hintz, A.-M. & Grünke, M. (2009). Einschätzungen von angehenden Lehrkräften für Sonder- und allgemeine Schulen zur Wirksamkeit von Interventionen für den Schriftspracherwerb bei lernschwachen Kindern. *Empirische Sonderpädagogik*, 1(1), 45-61.
- Holle, K. (2006). Flüssiges und phrasiertes Lesen (fluency): Lesetheoretische Grundlagen und unterrichtspraktische Hinweise. In S. Weinold (Hrsg.), *Schriftspracherwerb empirisch. Konzepte – Diagnostik – Entwicklung* (S. 87-119). Baltmannsweiler: Schneider Hohengehren.
- Homburg, C. & Giering, A. (1996). Konzeptualisierung und Operationalisierung komplexer Konstrukte. Ein Leitfaden für die Marketingforschung. *Marketing: Zeitschrift für Forschung und Praxis*, 18(1), 3-24.
- Hosenfeld, I. & Zimmer-Müller, M. (2014). *VERA. Vergleichsarbeiten in Grundschulen. Ein Projekt des Zentrums für Empirische Pädagogische Forschung (zepf), Universität Koblenz-Landau*. Verfügbar unter: <http://www.uni-landau.de/vera/> [Zugriff: Juni 2020].
- Huemer, S. M., Pointner, A., Schöfl, M. & Landerl, K. (2019). *Evidenzbasierte LRS-Förderung. Bericht über die wissenschaftlich überprüfte Wirksamkeit von Programmen und Komponenten, die in der LSR-Förderung zum Einsatz kommen*. [Broschüre im Auftrag des Bundesministeriums für Unterricht, Kunst und Kultur]. Verfügbar unter: www.schulpsychologie.at [Zugriff: Juni 2020].
- Hurrelmann, B. (1994). Leseförderung. Basisartikel. *Praxis Deutsch*, 21(127), 17-25.
- Hurrelmann, B. (2004). Sozialisation der Lesekompetenz. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (S. 37-60). Wiesbaden: VS Verlag für Sozialwissenschaften.

- Hurrelmann, B. (2011). Modelle und Merkmale der Lesekompetenz. In A. Bertschi- Kaufmann (Hrsg.), *Lesekompetenz Leseleistung Leseförderung. Grundlagen, Modelle und Materialien* (4. Aufl., S. 18-28). Zug: Klett, Kallmeyer.
- Husfeld, V. & Lindauer, T. (2009). Kompetenzen beschreiben und messen. Eine Problematisierung selbstverständlicher Begriffe. In A. Bertschi- Kaufmann & C. Rosebrock (Hrsg.), *Literalität. Bildungsaufgabe und Forschungsfeld* (S. 137-150). Weinheim: Juventa.
- Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E.-M. et al. (2017). *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- IBM Corporation. (2017). *IBM SPSS Statistics for Windows (Version 25)*. Armonk, New York: IBM Corporation.
- Inckemann, E. (2008). Förderdiagnostische Kompetenzen von Grundschullehrerinnen im schriftsprachlichen Anfangsunterricht. *Zeitschrift für Grundschulforschung*, 1(2), 99-115.
- Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der pädagogischen Diagnostik* (6). Weinheim, Basel: Beltz.
- Ise, E., Engel, R. R. & Schulte-Körne, G. (2012). Was hilft bei der Lese-Rechtschreibstörung? Ergebnisse einer Metaanalyse zur Wirksamkeit deutschsprachiger Förderansätze. *Kindheit und Entwicklung*, 21(2), 122-136.
- Isler, D. & Leemann, R. J. (2008). Literalität – wirksame Lese- und Schreibförderung im Unterricht. In U. Moser & J. Hollenweger (Hrsg.), *Drei Jahre danach. Lesen, Wortschatz und soziale Kompetenz am Ende der dritten Klasse* (S. 243-294). Oberentfelden: Sauerländer.
- Ite, N. (2015). Filmvignetten zur Einschätzung sprachförderrelevanten Wissens vor frühpädagogischen Fachpersonen. In C. Bräuer & D. Wieser (Hrsg.), *Lehrende im Blick. Empirische Lehrerforschung in der Deutschdidaktik* (S. 301-319). Wiesbaden: Springer VS.
- Jagemann, S. (2016). "Hör mal genau hin <Tru -He." – Wie verstehen und erklären angehende Lehrer/innen das silbenintitiale <h>? In H. Zimmermann & A. Peyer (Hrsg.), *Wissen und Normen – Facetten professioneller Kompetenz von Deutschlehrkräften* (S. 221-246). Frankfurt am Main: Peter Lang.
- Jandl, S. (2016). *Mathematikspezifisches Wissen von Sonderpädagoginnen und Sonderpädagogen. Eine empirische Studie zur Entwicklung, Evaluation und dem Einsatz eines Befragungsinstruments*. [Dissertation Universität Zürich]. Verfügbar unter: <https://www.zora.uzh.ch/id/eprint/133521/1/Jandl.pdf> [Zugriff: Juni 2020].
- Jandl, S. & Moser Opitz, E. (2017). Mathematische Förderung von Kindern mit intellektueller Beeinträchtigung. Über welches fachspezifische professionelle Wissen verfügen Sonderschullehrkräfte? *Sonderpädagogische Förderung heute*, 62(2), 195-208.
- Jansen, H., Mannhaupt, G., Marx, H. & Skowronek, H. (2002). *BISC. Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten. Manual* (2. überarbeitete Aufl.). Göttingen: Hogrefe.
- Joller-Graf, K. & Sturny-Bossart, G. (2010). Welche Kompetenzen sollen Schulische Heilpädagoginnen und Heilpädagogen in ihrer Ausbildung erwerben? *Schweizerische Zeitschrift für Heilpädagogik*, 16(1), 8-16.
- Juska-Bacher, B., Beckert, C., Stalder, U. & Schneider, H. (2016). Die Bedeutung des Wortschatzes für basale Lesekompetenzen. *Didaktik Deutsch. Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur*, 21(40), 20-40.
- Kaiser, G. & Busse, A. (2015). Wissen und Fähigkeiten in Fachdidaktik und Pädagogik: Zur Natur der professionellen Kompetenz von Lehrkräften. *Zeitschrift für Pädagogik*, 61(3), 328-344.
- Kämper-van den Boogaart, M. (2010). Lehrerkonzepte und Lehrerkompetenz für den Lese- und Literaturunterricht. In M. Kämper- van den Boogaart & K. H. Spinner (Hrsg.), *Deutschunterricht in Theorie und Praxis. (DTP). Lese- und Literaturunterricht. (Teil 2). Kompetenzen und Unterrichtsziele. Methoden und Unterrichtsmaterialien. Gegenwärtiger Stand der empirischen Unterrichtsforschung* (S. 104-136). Baltmannsweiler: Schneider-Verlag Hohengehren.
- Kamzela, K. (2015). Die Perspektive von Lehrenden auf basale Lesefähigkeiten zu Beginn der Sekundarstufe 1. In C. Bräuer & D. Wieser (Hrsg.), *Lehrende im Blick: Empirische Lehrerforschung in der Deutschdidaktik* (S. 49-70). Wiesbaden: Springer VS.

- Karing, C., Matthäi, J. & Artelt, C. (2011). Genauigkeit von Lehrerurteilen über die Lesekompetenz ihrer Schülerinnen und Schüler in der Sekundarstufe I – Eine Frage der Spezifität? *Zeitschrift für pädagogische Psychologie*, 25(3), 159-172.
- Karing, C., Pfost, M. & Artelt, C. (2011). Hängt die diagnostische Kompetenz von Sekundarstufenlehrkräften mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz ihrer Schülerinnen und Schüler zusammen? *Journal for educational research online*, 3(2), 119-147.
- Karst, K. (2017). Akkurate Urteile – die Ansätze von Schrader (1989) und McElvany et al. (2009). In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (S. 21-29). Münster: Waxmann.
- Kemmler, L. (1976). *Schulerfolg und Schulversagen. Eine Längsschnittuntersuchung vom ersten bis zum fünfzehnten Schulbesuchsjahr*. Göttingen: Hogrefe.
- Kernen, N. & Riss, M. (2012). *Textschwierigkeiten in Lehrmitteln für den naturwissenschaftlichen Unterricht in der Sekundarstufe I. Eine Analyse von der Pädagogischen Hochschule FHNW Zentrum Lesen*. Verfügbar unter: www.ilz.ch/cms/downloads/2012_Textschwierigkeiten.pdf, [Zugriff: Juni 2020].
- Kirby, J. R., Georgiou, G. K., Martinussen, R., Parrila, R., Bowers, P. & Landerl, K. (2010). Naming Speed and Reading: From Prediction to Instruction. *Reading Research Quarterly*, 45(3), 341-362.
- Kirchhoff, P. (2017). Falko-E: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. Mulder (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik* (S. 113-152). Münster: Waxmann.
- Kirschhock, E.-M. (2004). *Entwicklung schriftsprachlicher Kompetenzen im Anfangsunterricht*. Bad Heilbrunn: Julius Klinkhardt.
- Kleickmann, T. & Anders, Y. (2011). Lernen an der Universität. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 305-315). Münster Waxmann.
- Klicpera, C., Ehgartner, M., Gasteiger-Klicpera, B. & Schabmann, A. (1993). Lesenlernen in den ersten beiden Klassen der Sonderschule: Entwicklung der Wortlesefähigkeit bei lernbehinderten Kindern in der Sonderschule und bei guten und schwachen Lesern in der Grundschule. *Heilpädagogische Forschung*, 19(3), 97-103.
- Klicpera, C. & Gasteiger-Klicpera, B. (1993). *Lesen und Schreiben. Entwicklung und Schwierigkeiten. Die Wiener Längsschnittuntersuchungen über die Entwicklung, den Verlauf und die Ursachen von Lese- und Schreibschwierigkeiten in der Pflichtschulzeit*. Bern: Huber.
- Klicpera, C., Graeven, M. & Schabmann, A. (2000). Lesen- und Schreibenlernen in der Volksschule: Fortschritte in den ersten beiden Volksschulklassen, Prädiktoren der Fortschritte im Kindergarten und deren Abhängigkeit von der Gestaltung des Unterrichts und der Unterstützung in der Familie. Wien: Abteilung für angewandte und klinische Psychologie.
- Klicpera, C., Schabmann, A. & Gasteiger-Klicpera, B. (2013). *Legasthenie – LRS. Modelle, Diagnose, Therapie und Förderung* (4. aktualisierte Aufl.). München: Reinhardt UTB.
- Klicpera, C., Schabmann, A., Gasteiger-Klicpera, B. & Schmidt, B. (2017). *Legasthenie – LRS : Modelle, Diagnose, Therapie und Förderung* (5. überarbeitete und erweiterte Aufl.). München: Reinhardt UTB.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M. et al. (2007). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise im Auftrag des Bundesministerium für Bildungsforschung*. Berlin: Bundesministerium für Bildung und Forschung (BMBF).
- Klieme, E. & Hartig, J. (2008). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. In M. Prenzel, I. Gogolin & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik* (S. 11-29). Wiesbaden: VS Verlag für Sozialwissenschaften.

- Klusmann, U. (2011). Individuelle Voraussetzungen von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. (S. 297-304). Münster: Waxmann.
- König, J. (2010). Lehrerprofessionalität. Konzepte und Ergebnisse der internationalen und deutschen Forschung am Beispiel fachübergreifender, pädagogischer Kompetenzen. In J. König & B. Hofmann (Hrsg.), *Professionalität von Lehrkräften. Was sollen Lehrkräfte im Lese- und Schreibunterricht wissen und können?* (S. 40-105). Berlin: Deutsche Gesellschaft für Lesen und Schreiben.
- König, J. (2012). Die Entwicklung von pädagogischem Unterrichtswissen. Theoretischer Rahmen, Testinstrument, Skalierung und Ergebnisse. In J. König & A. Seifert (Hrsg.), *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerbildung* (S. 141-182). Münster: Waxmann.
- König, J. (2015). Kontextualisierte Erfassung von Lehrerkompetenz. *Zeitschrift für Pädagogik*, 3(61), 305-309.
- König, J. & Blömeke, S. (2010). Messung des pädagogischen Wissens: Theoretischer Rahmen und Teststruktur. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich* (S. 253-273). Münster: Waxmann.
- König, J., Doll, J., Buchholtz, N., Förster, S., Kaspar, K., Rühl, A.-M. et al. (2018). Pädagogisches Wissen versus fachdidaktisches Wissen? Struktur des professionellen Wissens bei angehenden Deutsch-, Englisch- und Mathematiklehrkräften im Studium. *Zeitschrift für Erziehungswissenschaft*, 21(3), 611-648.
- König, J. & Hofmann, B. (2010). Forschung zur Professionalität von Deutschlehrkräften. Aktuelle Entwicklungen und Desiderate. In J. König & B. Hofmann (Hrsg.), *Professionalität von Lehrkräften. Was sollen Lehrkräfte im Lese- und Schreibunterricht wissen und können?* (S. 7-20). Berlin: Deutsche Gesellschaft für Lesen und Schreiben.
- König, J. & Seifert, A. (2012). *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerbildung*. Münster: Waxmann.
- Konsortium-PISA. (2009). *PISA. Data Analysis Manual. SPSS Second Edition*: Organisation for Economic Co-operation and Development (OECD).
- Konsortium PISA. (2018). *PISA 2015: Schülerinnen und Schüler der Schweiz im internationalen Vergleich*. Bern, Genf: SKFI/EDK und Konsortium PISA.ch.
- Krauss, S., Blum, W., Neubrand, M., Baumert, J., Kunter, M., Besser, M. et al. (2011). Konzeptualisierung und Testkonstruktion zum fachbezogenen Professionswissen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 135-161). Münster: Waxmann.
- Krauss, S. & Bruckmaier, G. (2014). Das Experten-Paradigma in der Forschung zum Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. Aufl., S. 241-261). Münster: Waxmann.
- Krauss, S., Lindl, A., Schilcher, A., Fricke, M., Göhring, A., Hofmann, B. et al. (2017a). *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik*. Münster: Waxmann.
- Krauss, S., Lindl, A., Schilcher, A. & Tepner, O. (2017b). Das Forschungsprojekt FALKO – ein einleitender Überblick. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. Mulder (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik* (S. 9-65). Münster: Waxmann.
- Krauss, S., Neubrand, M., Blum, W., Baumert, J., Brunner, M. & Kunter, M. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie. *Journal für Mathematik-Didaktik*, 29(3-4), 233-258.

- Krebs, R. (2004). *Anleitung zur Herstellung von MC-Fragen und MC-Prüfungen für die ärztliche Ausbildung*. Verfügbar unter: <https://www.elearning.uni-mainz.de/files/2014/03/Anleitung-zur-Herstellung-von-MC-Fragen-und-MC-Pr%C3%BCfungen-f%C3%BCr-die-%C3%A4rztliche-Ausbildung.pdf> [Zugriff: Juli 2016].
- Kruse, G. (2006). *Lesekompetenz entwickeln – Lesestrategien trainieren*. Pädagogische Hochschule der FHNW. Verfügbar unter: https://web0.fhnw.ch/plattformen/zl/rs_10_2006_lesekompetenz-entwickeln-lesestrategien-trainieren/ [Zugriff: Juni 2020].
- Kruse, G. (2011). Das Lesen trainieren: Zu Konzepten von Leseunterricht und Leseübung. In A. Bertschi- Kaufmann (Hrsg.), *Lesekompetenz Leseleistung Leseförderung. Grundlagen, Modelle, Materialien* (4. Aufl., S. 176-188). Zug: Klett, Kallmeyer.
- Kruse, G., Rickli, U., Riss, M. & Sommer, T. (2010). *Lesen. Das Training. Teil 1 Unterstufe. Kommentar für Lehrerinnen und Lehrer*. Bern: Schulverlag plus.
- Kruth, I. & Thul, I. (2003). Kriterien zur Begriffs- und Bildauswahl bei der Anpassung von Anlauttabellen für Schüler mit Spezifischen Spracherwerbsstörungen. *Die Sprachheilarbeit*, 48(4), 157-163.
- Kulturministerkonferenz [KMK]. (2005). *Bildungsstandards im Fach Deutsch für den Primarbereich*. Neuweid: Luchterhand.
- Kulturministerkonferenz [KMK]. (2014). *Standards für die Lehrerbildung: Bildungswissenschaften (Beschluss der Kultusministerkonferenz vom 16.12.2004 in der Fassung vom 12.06.2014)*. Verfügbar unter: https://www.akkreditierungsrat.de/fileadmin/Seiteninhalte/KMK/Vorgaben/KMK_Lehrerbildung_Standards_Bildungswissenschaften_aktuell.pdf [Zugriff: Juni 2020].
- Kulturministerkonferenz [KMK]. (2019). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung (Beschluss der Kultusministerkonferenz vom 16.10.2008 in der Fassung vom 16.05.2019)*. Verfügbar unter: <https://www.bzl.uni-bonn.de/dokumente/fachprofile-lehrerbildung.pdf> [Zugriff: Juni 2020].
- Kunter, M. & Baumert, J. (2011). Das COACTIV-Forschungsprogramm zur Untersuchung professioneller Kompetenz von Lehrkräften – Zusammenfassung und Diskussion. In J. Baumert, M. Runter, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 345-366). Münster: Waxmann.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (2011a). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Kunter, M., Kleickmann, T., Klusmann, U. & Richter, D. (2011b). Die Entwicklung professioneller Kompetenz von Lehrkräften. In J. Baumert, M. Runter, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 55-68). Münster: Waxmann.
- Kunter, M. & Klusmann, U. (2010). Kompetenzmessung bei Lehrkräften – Methodische Herausforderungen (Measuring Teachers' Competence – Methodological Challenges). *Unterrichtswissenschaft*, 38(1), 68-86.
- Küspert, P. & Schneider, W. (2008). *Hören, lauschen, lernen. Sprachspiele für Kinder im Vorschulalter. Würzburger Trainingsprogramm zur Vorbereitung auf den Erwerb der Schriftsprache. Anleitung und Arbeitsmaterial*. (6. Aufl.). Göttingen: Vandenhoeck & Ruprecht.
- Landerl, K., Freudenthaler, H. H., Heene, M., De Jong, P. F., Desrochers, A., Manolitsis, G. et al. (2019). Phonological Awareness and Rapid Automatized Naming as Longitudinal Predictors of Reading in Five Alphabetic Orthographies with Varying Degrees of Consistency. *Scientific Studies of Reading*, 23(3), 220–234.
- Landerl, K. & Wimmer, H. (2008). Development of Word Reading Fluency and Spelling in a Consistent Orthography: An 8-Year Follow-Up. *Journal of Educational Psychology*, 100(1), 150-161.
- Langer, I., Schulz von Thun, F. & Tausch, R. (2011). *Sich verständlich ausdrücken* (9. Aufl.). München: Reinhardt.

- Lehmann, R. H., Peek, R. & Poerschke, J. (2006). *HAMLET 3-4: Hamburger Lesetest für 3. und 4. Klassen* (2. überarbeitete Aufl.). Göttingen: Hogrefe.
- Lenhard, A. & Lenhard, W. (2017). Diagnoseverfahren zur Erfassung schriftsprachlicher Leistungen. In M. Philipp (Hrsg.), *Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben* (S. 174-198). Weinheim: Beltz, Juventa.
- Lenhard, W. (2019). *Leseverständnis und Lesekompetenz: Grundlagen – Diagnostik – Förderung* (2. aktualisierte Aufl.). Stuttgart: Kohlhammer.
- Lenhard, W. & Artelt, C. (2009). Komponenten des Leseverständnisses. In W. Lehnhard & W. Schneider (Hrsg.), *Diagnostik und Förderung des Leseverständnisses* (4. unveränderte Aufl., S. 1-18). Göttingen: Hogrefe.
- Lenhard, W., Lenhard, A. & Schneider, W. (2020). *ELFE II: Ein Leseverständnistest für Erst- bis Siebtklässler – Version II* (4). Göttingen: Hogrefe.
- Lenhard, W., Schneider, W., Lenhard, A. & Lippard, J. (2006). *ELFE 1-6. Ein Leseverständnistest für Erst- bis Sechstklässler*. Göttingen: Hogrefe.
- Lessing-Sattari, M. & Wieser, D. (2018). Lehrkräfte. Systematisierung aktueller empirischer Studien, ihrer Gegenstandsbereiche und Forschungsansätze. In J. Boelmann (Hrsg.), *Empirische Forschung in der Deutschdidaktik. Band 3: Forschungsfelder der Deutschdidaktik* (S. 41-55). Baltmannsweiler: Schneider Verlag Hohengehren.
- Liancre, J. M. (2019). *Reliability and separation of measures*. Verfügbar unter: <https://www.winsteps.com/winman/reliability.htm> [Zugriff: April 2019].
- Liebers, K. & Prengel, A. (2010). *ILeA 1. Individuelle Lernstandsanalysen. Lehrerheft Deutsch/Mathematik* (6. überarbeitete Aufl.). Ludwigsfelde-Struveshof: Landesinstitut für Schule und Medien Berlin-Brandenburg (LISUM).
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.
- Lihuan, S. C. & Anthony, A. D. (2018). Examining Sources of Gender DIF in Mathematics Knowledge of Future Teachers Using Cross-Classfield IRT Models. In M. T. Totto, M. C. Rodriguez, W. M. Smith, M. D. Reckase & K. Bankov (Hrsg.), *Exploring the Mathematical Education of Teachers Using TEDS-M Data* (S. 543-562). Cham: Springer.
- Lindauer, T. & Schneider, H. (2011). Lesekompetenz ermitteln: Aufgaben im Unterricht. In A. Bertschi- Kaufmann (Hrsg.), *Lesekompetenz Leseleistung Leseförderung. Grundlagen, Modelle und Materialien* (4. Aufl., S. 109-125). Zug: Klett, Kallmeyer.
- Lindl, A. & Krauss, S. (2017). Transdisziplinäre Perspektive auf domänenspezifische Lehrerkompetenz. Eine Metaanalyse zentraler Resultate des Forschungsprojekts FALKO. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. Mulder (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik* (S. 381-438). Münster: Waxmann.
- Lindmeier, A. (2011). *Modeling and measuring knowledge and competencies of teachers. A threefold domain-specific structure model for mathematics*. Münster: Waxmann.
- Lindmeier, A. (2013). Video-vignettenbasierte standardisierte Erhebung von Lehrerkognitionen. In U. Riegel & K. Macha (Hrsg.), *Videobasierte Kompetenzforschung in den Fachdidaktiken. Fachdidaktische Forschung (Band 4)* (S. 45-61). Münster: Waxmann.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenz für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern* (S. 47-70). Weinheim: Beltz.
- Lombard, M., Snyder-Duch, J. & Bracken, C. (2002). Content Analysis in Mass Communication. Assessment and Reporting of Intercoder Reliability. *Human Communication Research*, 28(4), 587–604.
- Löwen, K., Baumert, J., Kunter, M., Krauss, S. & Brunner, M. (2011). Methodische Grundlagen des Forschungsprogramms. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 69-84). Münster: Waxmann.
- Lüdtke, O. & Robitzsch, A. (2017). Eine Einführung in die Plausible-Values-Technik für die psychologische Forschung. *Diagnostica*, 63(3), 193-205.

- Maag Merki, K. & Werner, S. (2014). Erfassung und Bewertung professioneller Kompetenz von Lehrpersonen. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. Aufl., S. 745-763). Münster: Waxmann.
- Mannhaupt, G. (2006). *MÜSC: Münsteraner Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten* (1. Aufl.). Berlin: Cornelsen.
- Martschinke, S., Kirschhock, E.-M. & Frank, A. (2018). *Der Rundgang durch Hörhausen Erhebungsverfahren zur phonologischen Bewusstheit Diagnose und Förderung im Schriftspracherwerb* (10. Aufl.). Donauwörth: Auer.
- Marx, H. (2007a). Theorien und Determinanten des Erwerbs der Schriftsprache. In H. Schöler & A. Welling (Hrsg.), *Sonderpädagogik der Sprache (Band 1)* (S. 92-147). Göttingen: Hogrefe.
- Marx, H. & Jungmann, T. (2000). Abhängigkeit der Entwicklung des Leseverstehens von Hörverstehen und grundlegenden Lesefertigkeiten im Grundschulalter: Eine Prüfung des Simple View of Reading-Ansatzes. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 32(2), 81-93.
- Marx, P. (2007b). *Lese- und Rechtschreiberwerb*. Paderborn: Schöningh.
- May, P. & Arntzen, H. (2003). *HLP: Hamburger Leseprobe HLP 1-4. Testverfahren zur Beobachtung der Leselernentwicklung in der Grundschule* (3. überarbeitete Aufl.). Hamburg: Selbstverlag.
- Mayer, A. (2011). *Test zur Erfassung der phonologischen Bewusstheit und der Benennungsgeschwindigkeit (TEPHOBE). Manual*. München: Reinhardt.
- Mayer, A. (2013). *Gezielte Förderung bei Lese- und Rechtschreibstörungen* (2. Aufl.). München: Reinhardt.
- Mayer, A. (2014). Früherkennung und Prävention von Schriftspracherwerbsstörungen im inklusiven Unterricht. In S. Sallat, M. Spreer & C. W. Glück (Hrsg.), *Sprache professionell fördern* (S. 390-401). Idstein: Schulz-Kirchner Verlag.
- Mayer, A. (2015). Kriterien zur Erstellung sprachlich optimierter Lesetexte für Kinder mit Sprachverständnisschwierigkeiten. *Praxis Sprache*, 60(4), 221-228.
- Mayer, A. (2016). *Lese-Rechtschreibstörungen (LRS)*. München: Reinhardt.
- Mayer, A. (2018a). Benennungsgeschwindigkeit und Lesen. *Forschung Sprache*, 6(1), 20-42.
- Mayer, A. (2018b). *Blitzschnelle Worterkennung (BliWo) : Grundlagen und Praxis* (3. überarbeitete Aufl.). Dortmund: Verlag Modernes Lernen.
- Mayring, P. (2003). *Qualitative Inhaltsanalyse. Grundlage und Techniken* (8. Aufl.). Weinheim: Beltz UTB.
- Mayringer, H. & Wimmer, H. (2014). *SLS 2-9. Salzburger Lese-Screening für die Schulstufen 2-9*. Bern: Hogrefe.
- McCutchen, D., Harry, D. R., Cunningham, A. E., Cox, S., Sidman, S. & Covill, A. E. (2002). Reading Teachers' Knowledge of Children's Literature and English Phonology. *Annals of Dyslexia*, 52, 207-228.
- McElvany, N., Becker, M. & Lüdtke, O. (2009). Die Bedeutung familiärer Merkmale für Lesekompetenz, Wortschatz, Lesemotivation und Leseverhalten. *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie*, 41(3), 121-131.
- McElvany, N., Kortenbruck, M. & Becker, M. (2008). Lesekompetenz und Lesemotivation. Entwicklung und Mediation des Zusammenhangs durch Leseverhalten. *Zeitschrift für pädagogische Psychologie*, 22(3-4), 207-219.
- McElvany, N. & Schneider, C. (2009). Förderung der Lesekompetenz. In W. Lehnhard & W. Schneider (Hrsg.), *Diagnostik und Förderung des Leseverständnisses* (S. 151-182). Göttingen: Hogrefe.
- Melzer, C. & Hillenbrand, C. (2013). Aufgaben sonderpädagogischer Lehrkräfte für die inklusive Bildung. Empirische Befunde internationaler Studien. *Zeitschrift für Heilpädagogik*, 64(5), 194-202.
- Melzer, C., Hillenbrand, C., Sprenger, D. & Hennemann, T. (2015). Aufgaben von Lehrkräften in inklusiven Bildungssystemen – Review internationaler Studien. *Erziehungswissenschaft* 26(51), 61-80.
- Messick, S. (1995). Standards of Validity and the Validity of Standards in Performance Assessment. *Educational Measurement: Issues and Practice*, 14(4), 5-8.
- Metze, W. (2002). *Tobi Fibelhandbuch*. Berlin: Cornelsen.

- Moats, L. C. (1994). The missing foundation in teacher education: Knowledge of the structure of spoken and written language. *Annals of Dyslexia*, 44(1), 81-102.
- Moats, L. C. & Foorman, B. R. (2003). Measuring Teachers' Content Knowledge of Language and Reading. *Annals of Dyslexia*, 53(1), 23-45.
- Moll, K. & Landerl, K. (2010). *SLRT-II. Lese- und Rechtschreibtest. Weiterentwicklung des Salzburger Lese- und Rechtschreibtests (SLRT)*. Bern: Huber.
- Moll, K. & Landerl, K. (2011). Lesedefizite und Rechtschreibdefizite – zwei Seiten derselben Medaille? In G. Schulte-Körne (Hrsg.), *Legasthenie und Dyskalkulie: Stärken erkennen – Stärken fördern* (S. 11-24). Bochum: Winkler.
- Moll, K. & Landerl, K. (2014). *SLRT-II. Lese- und Rechtschreibtest. Weiterentwicklung des Salzburger Lese- und Rechtschreibtests (SLRT)* (2. korrigierte Aufl.). Bern: Huber.
- Moosbrugger, H. (2012a). Item-Response-Theorie (IRT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. aktualisierte und überarbeitete Aufl., S. 227-292). Heidelberg: Springer.
- Moosbrugger, H. (2012b). Klassische Testtheorie (KTT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. aktualisierte und überarbeitete Aufl., S. 103-117). Heidelberg: Springer.
- Moosbrugger, H. & Kelava, A. (2012a). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. aktualisierte und überarbeitete Aufl., S. 7-26). Heidelberg: Springer.
- Moosbrugger, H. & Kelava, A. (2012b). *Testtheorie und Fragebogenkonstruktion* (2. aktualisierte und überarbeitete Aufl.). Heidelberg: Springer.
- Moosbrugger, H. & Kelava, A. (2020). Qualitätsanforderungen an Tests und Fragebogen („Gütekriterien“). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. vollständig neu bearbeitete, erweiterte und aktualisierte Aufl., S. 13-38). Berlin: Springer.
- Moosbrugger, H. & Schermelleh-Engel, K. (2012). Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. aktualisierte und überarbeitete Aufl., S. 325-343). Heidelberg: Springer.
- Moser Opitz, E. & Nührenbörger, M. (2015). Diagnostik und Leistungsbeurteilung. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 491-512). Berlin: Springer.
- Moser, V. & Kropp, A. (2014). *Abschlussbericht: „Kompetenzen in inklusiven setting“ (KIS) – Vorarbeiten zu einem Kompetenzstrukturmodell sonderpädagogischer Lehrkräfte*. [Abschlussbericht Humboldt-Universität zu Berlin]. Verfügbar unter: https://www.reha.huberlin.de/lehrgebiete/arp/forschung/moser-kropp_kis_abschlussbericht_2014.03.pdf [Zugriff: August 2019].
- Müller, B. (2015). *Förderung des Leseverständnisses durch Lesestrategien: Eine Interventionsstudie in der Grundschule*. [Dissertation Universität Kassel]. Verfügbar unter: <https://nbn-resolving.org/urn:nbn:de:hebis:34-2015012347267>; <http://d-nb.info/1065916051/34>; <https://kobra.bibliothek.uni-kassel.de/handle/urn:nbn:de:hebis:34-2015012347267> [Zugriff: Juni 2020].
- Müller, B., Krizan, A., Hecht, T., Richter, T. & Ennemoser, M. (2013). Leseflüssigkeit im Grundschulalter: Entwicklungsverlauf und Effekte systematischer Leseförderung. *Lernen und Lernstörungen*, 2(3), 131-146.
- Müller, M. & Buchs, H. (2014). *Intercoderreliabilität der Vercodung von Stellenausschreibungen. Stellenmarktmonitor Schweiz*. [Arbeitspapier der Universität Zürich]. Verfügbar unter: http://www.stellenmarktmonitor.uzh.ch/dam/jcr:00000000-4e80-02a5-0000-00000e31f0cf/Vercodungsreliabilitaet_150914_2.pdf [Zugriff: Juni 2017].
- Mullis, I. V. S., Martin, M. O., Foy, P. & Drucker, K. T. (2012). *PIRLS 2011: International Results in Reading*. Chestnut Hill, MA, USA: TIMSS & PIRLS International Study Center.
- Muthén, L. K. & Muthén, B. O. (1998-2012). *Mplus User's Guide* (7. Ed.). Los Angeles, CA: Muthén & Muthén.

- Muthén, L. K. & Muthén, B. O. (2015). *M Plus Computer Software (Version 7.31)*. Los Angeles, CA: Muthén & Muthén.
- National Reading Panel. (2000). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. Verfügbar unter: <http://www.nichd.nih.gov/publications/pubs/nrp/documents/report.pdf> [Zugriff: Juni 2020].
- Naumann, J., Artelt, C., Schneider, W. & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009 Bilanz nach einem Jahrzehnt* (S. 23-71). Münster: Waxmann.
- Neumann, K. (2014). Rasch-Analyse naturwissenschaftsbezogener Leistungstests. In D. Krüger, I. Prachmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 355-370). Heidelberg: Springer Spektrum.
- Neuweg, G. H. (2014). Das Wissen der Wissensvermittler. Problemstellungen, Befunde und Perspektiven der Forschung zum Lehrerwissen. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. Aufl., S. 583-614). Münster: Waxmann.
- Neuweg, G. H. (2015). Kontextualisierte Kompetenzmessung. Eine Bilanz zu aktuellen Konzeptionen und forschungsmethodischen Zugängen. *Zeitschrift für Pädagogik*, 61(3), 377-383.
- Niedermann, A. & Hartmann, E. (2006). Leseflüssigkeit als wichtiger Zielbereich der Förderung leseschwacher Kinder. Theoretische Grundlagen, methodische Ansätze und Anwendungsmöglichkeiten von Pattern Books. *MitSprache: Fachzeitschrift für Sprachheilpädagogik*, 38(1), 5-22.
- Niedermann, A. & Sassenroth, M. (2002). *Lesestufen. Ein Instrument zur Feststellung und Förderung der Leseentwicklung. Kommentar und Auswertungsbogen zum Bilderbuch "Dani hat Geburtstag"*. Zug: Klett, Balmer.
- Niedermann, A. & Sassenroth, M. (2017). *Lesestufen: Ein Instrument zur Feststellung und Förderung der Leseentwicklung* (7. Aufl.). Hamburg: Persen.
- Niermann, A. (2017). *Professionswissen von Lehrerinnen und Lehrern des Mathematik- und Sachunterrichts: "... man muss schon von der Sache wissen."*. Bad Heilbrunn: Klinkhardt Forschung.
- Nitko, A. J. & Brookhart, S. M. (2011). *Educational Assessment of Students* (6. Ed.). London: Pearson.
- Nix, D. (2010). Förderung der Lesekompetenz. In M. Kämper- van den Boogaart & K. H. Spinner (Hrsg.), *Deutschunterricht in Theorie und Praxis. (DTP). Lese- und Literaturunterricht (Teil 2). Kompetenzen und Unterrichtsziele. Methoden und Unterrichtsmaterialien. Gegenwärtiger Stand der empirischen Unterrichtsforschung* (S. 139-189). Baltmannsweiler: Schneider-Verlag Hohengehren.
- Nix, D. (2011). *Förderung der Leseflüssigkeit: Theoretische Fundierung und empirische Überprüfung eines kooperativen Lautlese Verfahrens im Deutschunterricht*. Weinheim: Juventa.
- Oser, F. (1997). Standards in der Lehrerbildung. Teil 1: Berufliche Kompetenzen, die hohen Qualitätsmerkmalen entsprechen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 15(1), 26-37.
- Oser, F. (2015). Einige Empfehlungen zu den in diesem Heft vorgestellten Analysen von Lehrerinnen- und Lehrerbildungsverläufen. Oder: Versteckte Normativitäten als Elemente der Kritik der empirischen Vernunft. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 33(1), 103-113.
- Oser, F., Curcio, G.-P. & Dügge, A. (2007). Kompetenzmessung in der Lehrerbildung als Notwendigkeit – Fragen und Zugänge. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 25(1), 14-26.
- Oser, F., Heinzer, S. & Salzmann, P. (2010). Die Messung der Qualität von professionellen Kompetenzprofilen von Lehrpersonen mit Hilfe der Einschätzung von Filmvignetten. Chancen und Grenzen des advokatorischen Ansatzes. *Unterrichtswissenschaft. Zeitschrift für Lernforschung*, 38(1), 5-28.
- Palincsar, A. S. & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117-175.

- Papadopoulos, T. C., Spanoudis, G. C. & Georgiou, G. K. (2016). How Is RAN Related to Reading Fluency? A Comprehensive Examination of the Prominent Theoretical Accounts. *Frontiers in Psychology*, 7, 1-15.
- Petermann, F. & Daseking, M. (2015). *ZLT-II: Zürcher Lesetest – II. Weiterentwicklung des Zürcher Lesetests (ZLT) von Maria Linder und Hans Grisseman* (3. überarbeitete Aufl. mit erweiterten Normen). Bern: Huber.
- Petermann, F. & Daseking, M. (2019). *ZLT-II: Zürcher Lesetest – II. Weiterentwicklung des Zürcher Lesetests (ZLT) von Maria Linder und Hans Grisseman* (4. überarbeitete Aufl.). Göttingen: Hogrefe.
- Pfost, M., Hattie, J., Dörfler, T. & Artelt, C. (2014). Individual Differences in Reading Development: A Review of 25 Years of Empirical Research on Matthew Effects in Reading. *Review of Educational Research*, 84(2), 203-244.
- Phelps, G. & Schilling, G. S. (2004). Developing Measures of Content Knowledge for Teaching Reading. *Elementary School Journal*, 105(1), 31-48.
- Philipp, M. (2012a). *Besser lesen und schreiben. Wie Schüler effektiver mit Sachtexten umgehen lernen*. Stuttgart: Kohlhammer.
- Philipp, M. (2012b). Lesestrategien – wirksame Werkzeuge für das Textverstehen. *Schweizerische Zeitschrift für Heilpädagogik*, 18(5), 40-45.
- Philipp, M. (2012c). *Motiviert lesen und schreiben. Dimensionen, Bedeutung, Förderung*. Seelze: Kallmeyer, Klett.
- Philipp, M. (2015). *Lesestrategien: Bedeutung, Formen und Vermittlung*. Weinheim: Beltz.
- Philipp, M. (2017). Kapitel 5: Entwicklung hierarchiehoher Prozesse. In M. Philipp (Hrsg.), *Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben* (S. 67-83). Weinheim: Beltz, Juventa.
- Piasta, S. B., McDonald Connor, C., Fishman, B. J. & Morrison, F. J. (2009). Teachers' Knowledge of Literacy Concepts, Classroom Practices, and Student Reading Growth. *Scientific Studies of Reading*, 13(3), 224-248.
- Pinnell, G. S. (1995). *Listening to Children Read Aloud: Data from NAEP's Integrated Reading Performance Record (IRPR) at Grade 4*. Washington, DC: National Center for Education Statistics.
- Pissarek, M. & Schilcher, A. (2015). Fachspezifische Lehrerkompetenzen im Fach Deutsch messen? – Modellierung und Konstruktvalidierung eines Erhebungsinstruments im Rahmen der Projektgruppe FALKO Regensburg. In C. Bräuer & D. Wieser (Hrsg.), *Lehrende im Blick. Empirische Lehrerforschung in der Deutschdidaktik* (S. 321-340). Wiesbaden: Springer VS.
- Pissarek, M. & Schilcher, A. (2017). FALKO-D: Die Untersuchung des Professionswissens von Deutschlehrenden. Entwicklung eines Messinstruments zur fachspezifischen Lehrerkompetenz und Ergebnisse zu dessen Validierung. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. Mulder (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik*. (S. 68-111). Münster: Waxmann.
- Pool Maag, S. & Moser Opitz, E. (2014). Inklusiver Unterricht – grundsätzliche Fragen und Ergebnisse einer explorativen Studie. *Empirische Sonderpädagogik*, 6(2), 133-149.
- Prost, R. (2014). *Fragebogen. Ein Arbeitsbuch* (4. erweiterte Aufl.). Wiesbaden: Springer VS.
- Ramseier, E. (2008). Validation of competence models for developing education standards: Methodological choices and their consequences. *Mesure et évaluation en éducation*, 31(2), 35-53.
- Ramseier, E., Labudde, P. & Adamina, M. (2011). Validierung des Kompetenzmodells HarmoS Naturwissenschaften. Fazite und Defizite. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 7-33.
- Ramseier, E., Moser, U., Moreau, J. & Antonietti, J.-P. (2008). HarmoS – Entwicklung von Bildungsstandards. Schlussbericht der Methodologiegruppe. Bern: Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK).
- Rasinski, T. V. (2003). *The Fluent Reader: Oral Reading Strategies for Building Word Recognition, Fluency, and Comprehension*. New York: Scholastic Professional Books.

- Rasinski, T. V. (2017). Readers Who Struggle: Why Many Struggle and a Modest Proposal for Improving Their Reading. *The Reading Teacher*, 70(5), 519-524.
- Reinisch, H. (2009). "Lehrprofessionalität" als theoretischer Term. Eine begriffssystematische Analyse. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität Bedingungen, Genese, Wirkungen und ihre Messung* (S. 33-43). Weinheim: Beltz.
- Reintjes, C., Keller, S., Jünger, S. & Düggeli, A. (2016). Aufgaben (in) der Ausbildung von Lehrerinnen und Lehrern. Theoretische Konzepte, Entwicklungs- und Forschungsperspektiven. In S. Keller & C. Reintjes (Hrsg.), *Aufgaben als Schlüssel zur Kompetenz. Didaktische Herausforderungen, wissenschaftliche Zugänge und empirische Befunde* (S. 429-447). Münster: Waxmann.
- Richter, D. & Lenski, A. (2015). Kompetenzorientierung im Unterricht aus der Perspektive von Lehrkräften und Schülerinnen und Schülern. *Zeitschrift für Pädagogik*, 61(5), 712-737.
- Richter, T. & Christmann, U. (2002). Lesekompetenz: Prozessebenen und interindividuelle Unterschiede. In N. Goeben & B. Hurrelmann (Hrsg.), *Lesekompetenz: Bedingungen, Dimensionen, Funktionen* (S. 25-58). Weinheim: Juventa.
- Richter, T., Isberner, M.-B., Naumann, J. & Kutzner, Y. (2012). Prozessbezogene Diagnostik von Lesefähigkeiten bei Grundschulkindern. *Zeitschrift für pädagogische Psychologie*, 26(4), 313-331.
- Richter, T. & Müller, B. (2017). Entwicklung hierarchieniedriger Prozesse. In M. Philipp (Hrsg.), *Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben* (S. 51-66). Weinheim: Beltz, Juventa.
- Richter, T., Naumann, J., Isberner, M.-B., Neeb, Y. & Knoepke, J. (2017). *ProDi-L: Prozessbezogene Diagnostik von Lesefähigkeiten im Grundschulalter (Computergestütztes Testverfahren)*. Göttingen: Hogrefe.
- Riegler, S. (2010). Auf die richtige Spur gesetzt. Das System der Buchstaben-Laut-Beziehung in einer Anlauttabelle. *Praxis Deutsch. Zeitschrift für den Deutschunterricht*, 37(221), 58-60.
- Riegler, S. & Wiprächtiger-Geppert, M. (2016). Konzeptneutral und unterrichtsnah. Wissen und Normen – Facetten professioneller Kompetenz von Lehrkräften. In H. Zimmermann & A. Peyer (Hrsg.), *Ein Instrument zur Erfassung des Professionswissens zu Orthographie und Orthographieerwerb* (S. 199-219). Frankfurt am Main: Peter Lang.
- Ritter, C. (2005). *Entwicklung und empirische Überprüfung eines Lesetrainings aus Silbenbasis*. [Dissertation der Humanwissenschaftlichen Fakultät Potsdam]. Verfügbar unter: https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/920/file/ritter_diss.pdf [Zugriff: Juni 2020].
- Ritter, C. & Scheerer-Neumann, G. (2009). *PotsBlitz – Potsdamer Lesetraining. Förderung der basalen Lesefähigkeiten. Manual*. Köln: ProLog.
- Rjosk, C., McElvany, N., Anders, Y. & Becker, M. (2011). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung der basalen Lesefähigkeit ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 58(2), 92-105.
- Röber, C. (2013). *Die Leistungen der Kinder beim Lesen- und Schreibenlernen : Grundlagen der Silbenanalytischen Methode: Ein Arbeitsbuch mit Übungsaufgaben* (3. Aufl.). Baltmannsweiler: Schneider Verlag Hohengehren.
- Röber, C. & Olfert, H. (2010). *Die Bedingungen für ein erfolgreiches Arbeiten mit Silben beim Lesen- und Schreibenlernen. Chancen und Grenzen der Konzepte der neuen Silbenfibeln* [Unveröffentlichte Stellungnahme der Deutschen Gesellschaft für Lesen und Schreiben /PH Freiburg]. Verfügbar unter: https://dgls.de/wp-content/uploads/2017/01/Stellungnahme_Silbenfibeln.pdf [Zugriff: Juni 2020].
- Rosebrock, C. (2012). Was ist Lesekompetenz, und wie kann sie gefördert werden? *Online-Plattform für Literalität / Leseforum.ch*. (3), 1-12.
- Rosebrock, C. & Nix, D. (2006). Forschungsüberblick: Leseflüssigkeit (Fluency) in der amerikanischen Leseforschung und -didaktik. *Didaktik Deutsch. Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur*, 20(1), 90-112.
- Rosebrock, C. & Nix, D. (2012). *Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung* (5. Aufl.). Baltmannsweiler: Schneider-Verlag Hohengehren.

- Rosebrock, C., Nix, D., Rieckmann, C. & Gold, A. (2011). *Lese flüssigkeit fördern. Lautleseverfahren für die Primar- und Sekundarstufe*. Seelze: Kallmeyer.
- Rosebrock, C., Rieckmann, C., Nix, D. & Gold, A. (2010). Förderung der Lese flüssigkeit bei leseschwachen Zwölfjährigen. *Didaktik Deutsch. Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur*, 16(28), 33-58.
- Rost, D. H. & Hartmann, A. (1992). Lesen, Hören, Verstehen. *Zeitschrift für Psychologie*, 200, 345-361.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2. vollständig überarbeitete und erweiterte Aufl.). Bern: Huber.
- Roters, B., Nold, G., Haudeck, H., Kessler, J.-U. & Stancel-Piatak, A. (2011). Professionelles Wissen von Studierenden des Lehramts Englisch. In S. Blömeke, A. Bremerich-Vos, H. Haudeck, G. Kaiser, G. Nold, K. Schwippert & H. Willenberg (Hrsg.), *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT* (S. 77-99). Münster: Waxmann.
- Rupp, G. & Bonholt, H. (2006). Lehr-Lern-Forschung als empirische Lese-/Literaturdidaktik?! In N. Groeben & B. Hurrelmann (Hrsg.), *Empirische Unterrichtsforschung in der Literatur- und Lesedidaktik: Ein Weiterbildungsprogramm* (S. 239-253). Weinheim: Juventa.
- Rutsch, J. (2016). *Entwicklung und Validierung eines Vignettentests zur Erfassung des fachdidaktischen Wissens im Leseunterricht bei angehenden Lehrkräften*. [Dissertation Pädagogische Hochschule Heidelberg]. Verfügbar unter: https://opus.ph-heidelberg.de/frontdoor/deliver/index/docId/215/file/Dissertation_Rutsch_Veroeffentlichung.pdf [Zugriff: Juni 2020].
- Rutsch, J., Rehm, M., Vogel, M. & Seidenfuß, M. (2018a). *Effektive Kompetenzdiagnose in der Lehrerbildung. Professionalisierungsprozesse angehender Lehrkräfte untersuchen*. Wiesbaden: Springer.
- Rutsch, J., Schmitt, M. & Dörfler, T. (2018b). Einfluss von Lehrerfahrung und spezifischen Lerngelegenheiten auf die Entwicklung des fachdidaktischen Wissens im Leseunterricht. In J. Rutsch, M. Rehm, M. Vogel, M. Seidenfuß & T. Dörfler (Hrsg.), *Effektive Kompetenzdiagnose in der Lehrerbildung Professionalisierungsprozesse angehender Lehrkräfte untersuchen* (S. 75-89). Wiesbaden: Springer.
- Rutsch, J., Seidenfuß, M., Vogel, M. & Rehm, M. (2017). Fachdidaktische Unterrichtsvignetten in Forschung und Lehre: Überblick über Forschungsarbeiten und Einsatzmöglichkeiten. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 35(3), 487-505.
- Rutsch, J., Vogel, M., Seidenfuß, M., Dörfler, T. & Rehm, M. (2018c). Professionalisierungsprozesse angehender Lehrkräfte untersuchen. In J. Rutsch, M. Rehm, M. Vogel, M. Seidenfuß & T. Dörfler (Hrsg.), *Effektive Kompetenzdiagnose in der Lehrerbildung* (S. 9-25). Wiesbaden: Springer.
- Sappok, C. & Fay, J. (2018). Prosodische Aspekte von Lese flüssigkeit messen. Evaluation einer Ratingprozedur mit Audioaufnahmen von DrittklässlerInnen. *Didaktik Deutsch. Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur*, 23(44), 61-83.
- Schabmann, A., Schmidt, B., Klicpera, C., Gasteiger-Klicpera, B. & Klingebiel, K. (2009). Does systematic reading instruction impede prediction of reading a shallow orthography? *Psychology Science Quarterly*, 51(3), 315-338.
- Scheerer-Neumann, G. (1981). The utilization of intraword structure in poor readers: Experimental evidence and a training program. *Psychological Research*, 43, 155-178.
- Scheerer-Neumann, G. (1997). Lesen und Leseschwierigkeiten. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (S. 279-235). Göttingen: Hogrefe.
- Scheerer-Neumann, G. (1998). Stufenmodelle des Schriftspracherwerbs – Wo stehen wir heute? In H. Balhorn (Hrsg.), *Schatzkiste Sprache 1: Von den Wegen der Kinder in die Schrift* (S. 14-46). Frankfurt am Main: Arbeitskreis Grundschule.
- Scheerer-Neumann, G. (2006). Das Lesen lernen (wollen). Wie das Eintauchen in die Welt der Buchstaben gelingt. *Grundschulunterricht*, 53(5), 16-20.
- Scheerer-Neumann, G. (2015). *Lese-Rechtschreib-Schwäche und Legasthenie. Grundlagen, Diagnostik und Förderung*. Stuttgart: Kohlhammer.
- Scheerer-Neumann, G. (2018). *Lese-Rechtschreib-Schwäche und Legasthenie: Grundlagen, Diagnostik und Förderung* (2. aktualisierte Aufl.).

- Scheerer-Neumann, G. & Hoffmann, C. D. (2005). Dimensionen der Lesekompetenz analysieren und fördern. In A. Sasse & R. Valtin (Hrsg.), *Lesen lehren* (S. 43-69). Berlin: Deutsche Gesellschaft für Lesen und Schreiben.
- Scherf, D. (2013). *Leseförderung aus Lehrersicht. Eine qualitativ-empirische Untersuchung professionellen Wissens*. Wiesbaden: Springer VS.
- Scherf, D. (2016). Was Lesedidaktiker sich wünschen (und was Lehrer tun): Lesedidaktische Vorstellungen professionellen Lehrerhandelns. In H. Zimmermann & A. Peyer (Hrsg.), *Wissen und Normen – Facetten professioneller Kompetenz von Lehrkräften* (S. 25-40). Frankfurt am Main: Peter Lang.
- Schermelleh-Engel, K. & Werner, C. (2012). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. aktualisierte und überarbeitete Aufl., S. 120-141). Berlin: Springer.
- Schindler, V., Moser Opitz, E., Cadonau-Bieler, M. & Ritterfeld, U. (2019). Überprüfung und Förderung des mathematischen Fachwortschatzes der Grundschulmathematik – eine empirische Studie. *Journal für Mathematik-Didaktik*, 40, 1-35.
- Schlicher, A. & Krauss, S. (2019). Anforderungen an die Lehrerausbildung aus fachdidaktischer Perspektive. In N. McElvany, W. Bos, H. G. Holtappels & A. Ohle-Peters (Hrsg.), *Bedingungen und Effekte von Lehrerbildung, Lehrkraftkompetenzen und Lehrkrafthandeln. Dortmunder Symposium der Empirischen Bildungsforschung* (S. 29-49). Münster: Waxmann.
- Schmidt, B. M. & Schabmann, A. (2010). "Es ist vorübergehend!" Lehrereinschätzungen über mögliche Lese-Rechtschreibprobleme – eine klassifikatorische Analyse. *Heilpädagogische Forschung*, 36(3), 106-115.
- Schmidt, B. M. & Schabmann, A. (2016). Wissen und Kompetenzwahrnehmung von Referendaren im Umgang mit LRS. *Heilpädagogische Forschung*, 42(2), 96-108.
- Schmidt, F. (2015). Den diagnostischen Blick schärfen – Vorstellungen und Orientierungen von Deutschlehrerinnen und Deutschlehrern zur Diagnose von Lesekompetenz. In C. Bräuer & D. Wieser (Hrsg.), *Lehrende im Blick. Empirische Lehrerforschung in der Deutschdidaktik* (S. 89-108). Wiesbaden: Springer VS.
- Schmidt, W., Blömeke, S., Tatto, M. T., Hsieh, F.-J., Cogan, L. S., Houang, R. T. et al. (2011). *Teacher Education Matters: A Study of Middle School Mathematics Teacher Preparation in Six Countries*. New York: Teachers College Press.
- Schmill, S. (2013). Wie verwenden Lehrkräfte Assessments zur Planung individualisierten Unterrichts im Bereich Lesen? Forschungsdesign und erste Ergebnisse einer qualitativen Fallstudie im Kontext eines Schulversuchs. *Didaktik Deutsch: Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur*, 19(35), 23-42.
- Schneider, H., Becker-Mrotzek, M., Sturm, A., Jambor-Fahlen, S., Neugebauer, U., Efinger, C. et al. (2013, 2013). *Expertise: Wirksame Sprachförderung*. [Bildungsdirektion des Kantons Zürich]. Verfügbar unter: <https://web0.fhnw.ch/plattformen/zi/expertise-wirksamkeit-von-sprachfoerderung/> [Zugriff: Juni 2020].
- Schneider, H. & Lindauer, T. (2011). Vorwort Rundschreiben Zentrum Lesen. *Rundschreiben Zentrum Lesen*(20), 1.
- Schneider, W. (2009). Diagnose basaler Lesekompetenzen in der Primar- und Sekundarstufe. In W. Lehnhard & W. Schneider (Hrsg.), *Diagnostik und Förderung des Leseverständnisses* (S. 45-64). Göttingen: Hogrefe.
- Schneider, W. (2017). *Lesen und Schreiben lernen. Wie erobern Kinder die Schriftsprache?* Berlin: Springer.
- Schneider, W., Blanke, I., Faust, V. & Küspert, P. (2011). *WLLP-R: Würzburger Leise Leseprobe – Revision. Ein Gruppentest für die Grundschule*. Göttingen: Hogrefe.
- Schnitzler, C. (2008). *Phonologische Bewusstheit und Schriftspracherwerb*. Stuttgart: Thieme.
- Schrader, F.-W. (2013). Diagnostische Kompetenzen von Lehrpersonen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 31(2), 154-165.
- Schrader, F.-W. (2014). Lehrer als Diagnostiker. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. Aufl., S. 865-882). Münster: Waxmann.
- Schröder-Lenzen, A. (2009). *Schriftspracherwerb und Unterricht. Bausteine professionellen Handlungswissens* (3. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.

- Schründer-Lenzen, A. (2013). *Schriftspracherwerb* (4. völlig überarbeitete Aufl.). Wiesbaden: Springer VS.
- Schründer-Lenzen, A. & Merkens, H. (2006). Differenzen schriftsprachlicher Kompetenzentwicklung bei Kindern mit und ohne Migrationshintergrund. In A. Schründer-Lenzen (Hrsg.), *Risikofaktoren kindlicher Entwicklung* (S. 15-44). Wiesbaden: VS Verlag für Sozialwissenschaften
- Schulte-Körne, G. (2014a). Aktuelle Entwicklungen zur Klassifikation und Definition der Lese-, der Rechtschreib- und der Lese-Rechtschreibstörung. In G.-K. Schulte & G. Thomé (Hrsg.), *LRS – Legasthenie: interdisziplinär* (S. 137-149). Oldenburg: isb Verlag.
- Schulte-Körne, G. (2014b). DSM-5 Kommentar. Spezifische Lernstörungen: Vom DSM-IV zum DSM-5. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 42(5), 369-374.
- Schulte-Körne, G. & Galuschka, K. (2019). *Lese-/Rechtschreibstörung (LRS)*. Göttingen: Hogrefe.
- Schulte, K., Nonte, S. & Schwippert, K. (2013). Die Überprüfung von Messinvarianz in international vergleichenden Schulleistungstudien am Beispiel der Studie PIRLS. *Zeitschrift für Bildungsforschung*, 3(99), 99-118.
- Schumacher, S. (2016). *Lehrerprofessionswissen im Kontext beschreibender Statistik Entwicklung und Aufbau des Testinstruments BeSt Teacher mit ausgewählten Analysen*. Wiesbaden: Springer Spektrum.
- Schwab, S. & Helm, C. (2015). Überprüfung von Messinvarianz mittels CFA und DIF-Analysen. *Empirische Sonderpädagogik*, 7(3), 175-193.
- Seemann, S., Dunekacke, S., Lindmeier, A., Heinze, A., Leuchter, M., Moser Opitz, E. et al. (2017). Anforderungsbezogene Modellierung und Erfassung domänenspezifischer professioneller Kompetenz frühpädagogischer Fachkräfte. In K. Ulrich & A. Kuzle (Hrsg.), *Beiträge zum Mathematikunterricht* (S. 1225-1228). Münster: WTM-Verlag.
- Seidenberg, M. S. & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review*, 96(4), 523-568.
- Seifert, A. (2015). Kompetenzforschung in den Fachdidaktiken auf der Grundlage von IRT-Modellen. In U. Riegel, S. Schubert, G. Siebert-Ott & K. Macha (Hrsg.), *Kompetenzmodellierung und Kompetenzmessung in den Fachdidaktiken* (S. 131-161). Münster: Waxmann.
- Seifert, A. & Schaper, N. (2012). Die Entwicklung von bildungswissenschaftlichem Wissen: Theoretischer Rahmen, Testinstrument und Ergebnisse. In J. König & A. Seifert (Hrsg.), *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerbildung* (S. 183-214). Münster: Waxmann.
- Seigneuric, A. & Ehrlich, M.-F. (2005). Contribution of Working Memory Capacity to Children's Reading Comprehension: A Longitudinal Investigation. *Reading and Writing*, 18(7), 617-656.
- Seuring, V. A. (2010). *Förderung des Leseverständnis mit Methoden des reziproken Lehrens. Effekt unterrichtsintegrierter Trainings für Schülerinnen und Schüler der 5. Klasse*. [Dissertation: Psychologie und Sportwissenschaft der Justus-Liebig-Universität Giessen]. Verfügbar unter: http://geb.uni-giessen.de/geb/volltexte/2010/7925/pdf/SeuringVanessa_2010_12_16.pdf [Zugriff: Juni 2020].
- Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15(2), 4-14.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57, 1-22.
- Shulman, L. S. (2004). *The Wisdom of Practice: Essays on Teaching, Learning, and Learning to Teach*. San Francisco: Jossey-Bass.
- Skowronek, H. & Marx, H. (1989). Die Bielefelder Längsschnittstudie zur Früherkennung von Risiken der Lese-Rechtschreibschwäche: Theoretischer Hintergrund und erste Befunde. *Heilpädagogische Forschung*, 15(1), 38-49.
- Sodogé, A. (2010). "Förderung ist Detektivarbeit". Was schulische Heilpädagoginnen und Heilpädagogen über die Förderung von Schülerinnen mit LRS denken. *Schweizerische Zeitschrift für Heilpädagogik*, 16(4), 21-28.
- Souvignier, E. (2009). Effektivität von Interventionen zur Verbesserung des Leseverständnisses. In W. Lenhard & W. Schneider (Hrsg.), *Diagnostik und Förderung des Leseverständnisses*. (S. 185-206). Göttingen: Hogrefe.

- Souvignier, E., Duzy, D. & Schneider, W. (2012). Effekte einer kombinierten Förderung des Sprachverstehens und der phonologischen Bewusstheit zur Prävention von Leserechtschreibschwierigkeiten. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete (VHN)*, 81(4), 282-298.
- Souvignier, E. & Förster, N. (2011). Effekte prozessorientierter Diagnostik auf die Entwicklung der Lesekompetenz leseschwacher Viertklässler. *Empirische Sonderpädagogik*(3), 243-255.
- Spinath, B. (2019). Anforderungen an die Lehrerbildung aus psychologischer Perspektive. In N. McElvany, W. Bos, H. G. Holtappels & A. Ohle-Peters (Hrsg.), *Bedingungen und Effekte von Lehrerbildung, Lehrkraftkompetenzen und Lehrkraft Handeln. Dortmunder Symposium der Empirischen Bildungsforschung* (S. 9-27). Münster: Waxmann.
- Stancel-Piatak, A., Buchholtz, C. & Schwippert, K. (2013). Anlage und Design der Studie TEDS-LT. In S. Blömeke, A. Bremerich-Vos, G. Kaiser, G. Nold, H. Haudeck, J.-U. Kessler & K. Schwippert (Hrsg.), *Professionelle Kompetenzen im Studienverlauf. Weitere Ergebnisse zur Deutsch-, Englisch- und Mathematiklehrerbildung aus TEDS-LT* (S. 25-46). Münster: Waxmann.
- Steck, A. (2010). "Ich mache zu irgend einem Stück so eine Art Leseverständnistest" – Über die Notwendigkeit einer systematischen Lehrerfortbildung im Bereich Leseverstehen. In J. König & B. Hofmann (Hrsg.), *Professionalität von Lehrkräften. Was sollen Lehrkräfte im Lese- und Schreibunterricht wissen und können?* (S. 214-226). Berlin: Deutsche Gesellschaft für Lesen und Schreiben.
- Stock, C., Marx, P. & Schneider, W. (2017). *BAKO 1-4. Basiskompetenzen für Leserechtschreibleistungen. Ein Test zur Erfassung der phonologischen Bewusstheit vom ersten bis vierten Grundschuljahr* (2. ergänzte und aktualisierte Aufl.). Göttingen: Hogrefe.
- Sturm, A. (2012). Wenn Lesen und Schreiben alles andere als selbstverständlich sind. *Schweizerische Zeitschrift für Heilpädagogik*, 18(5), 5-10.
- Sturm, A. (2016). Beurteilen und Kommentieren von Texten als fachdidaktisches Wissen. *Leseräume. Zeitschrift für Literalität in der Schule und Forschung*, 3(3), 115-132.
- Sturm, A., Schneider, H., Lindauer, N. & Sommer, T. (2016). Schreibbezogenes Fachwissen bei Lehrpersonen im ersten Berufsjahr. In M. Krelle & W. Senn (Hrsg.), *Qualität von Deutschunterricht* (S. 139-161). Stuttgart: Klett.
- Tacke, G. (2005). Evaluation eines Lesetrainings zur Förderung lese-rechtschreibschwacher Grundschüler der zweiten Klasse. *Psychologie in Erziehung und Unterricht*, 52 (3), 198-209
- Tacke, G. (2008). Frühe Leseförderung zur Vorbeugung von Leseschwäche. In J. Borchert, B. Hartke & P. Jogschies (Hrsg.), *Frühe Förderung entwicklungsauffälliger Kinder und Jugendlicher* (S. 151-163). Stuttgart: Kohlhammer.
- Tacke, G. (2011). Ein umfassendes Konzept zur schulischen und häuslichen Leserechtschreibförderung von Klasse 1 bis in die Sekundarstufe. In G. Schulte-Körne (Hrsg.), *Legasthenie und Dyskalkulie: Stärken erkennen – Stärken fördern* (S. 135-163). Bochum: Winkler.
- Tennant, A. & Pallant, J. F. (2007). DIF matters: A practical approach to test if Differential Item Functioning makes a difference. *Rasch Measurement Transactions*, 20(4), 1082-1084.
- Tenorth, H.-E. (2006). Professionalität im Lehrerberuf. Ratlosigkeit der Theorie, gelingende Praxis. *Zeitschrift für Erziehungswissenschaft*, 9(4), 580-597.
- Terhart, E. (2005). Standards in der Lehrerbildung – ein Kommentar. *Zeitschrift für Pädagogik*, 51(2), 275-279.
- Terhart, E. (2012). Wie wirkt Lehrerbildung? Forschungsprobleme und Gestaltungsfragen. *Zeitschrift für Bildungsforschung*, 2(1), 3-21.
- Therrien, W. J. (2004). Fluency and Comprehension Gains as a Result of Repeated Reading: A Meta-Analysis. *Remedial and Special Education*, 25(4), 252-261.
- Thomé, G. (2017). *ABC und andere Irrtümer über Orthographie, Rechtschreiben, LRS/Legasthenie* (4. erweiterte Aufl.). Oldenburg: Institut für sprachliche Bildung.
- Tiedemann, J. & Billmann-Mahecha, E. (2007). Macht das Fachstudium einen Unterschied? Zur Rolle der Lehrerexpertise für Lernerfolg und Motivation in der Grundschule. *Zeitschrift für Pädagogik*, 53(1), 58-73.

- Tillmann, K.-J. (2014). Konzepte der Forschung zum Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. überarbeitete und erweiterte Aufl., S. 308-316). Münster: Waxmann.
- Topsch, W. (2005). *Grundkompetenz Schriftspracherwerb. Methoden und handlungsorientierte Praxisanregungen* (2. überarbeitete und erweiterte Aufl.). Weinheim: Beltz.
- Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S. & Hecht, S. (1997). Contributions of Phonological Awareness and Rapid Automatic Naming Ability to the Growth of Word-Reading Skills in Second-to Fifth-Grade Children. *Scientific Studies of Reading*, 1(2), 161-185.
- Torppa, M., Tolvanen, A., Poikkeus, A.-M., Eklund, K., Lerkkanen, M.-K., Leskinen, E. et al. (2007). Reading Development Subtypes and Their Early Characteristics. *Annals of Dyslexia*, 57(1), 3-32.
- Trenk-Hinterberger, I., Nix, D., Rieckmann, C., Rosebrock, C. & Gold, A. (2007). Förderung der Leseflüssigkeit bei schwachen Leser(inne)n in der sechsten Jahrgangsstufe: Erste Ergebnisse einer Interventionsstudie. In B. Hofmann & R. Valtin (Hrsg.), *Checkpoint Literacy. Tagungsband zum 15. Europäischen Lesekongress* Berlin: Deutsche Gesellschaft für Lesen und Schreiben.
- Tresch, S. (2007). *Potential Leistungstest. Wie Lehrerinnen und Lehrer Ergebnismeldungen zur Sicherung und Steigerung ihrer Unterrichtsqualität nutzen*. Bern: h.e.p. Verlag.
- Tristán, A. (2006). An Adjustment for Sample Size in DIF Analysis. *Rasch Measurement Transactions*, 20(3), 88-92.
- Valtin, R. (2010). Phonologische Bewusstheit – eine notwendige Voraussetzung beim Lesen- und Schreibenlernen? *Online-Plattform für Literalität / Leseforum.ch*(2), 1-10.
- Vauras, M., Kinnunen, R. & Kuusela, L. (1994). Development of Text-Processing Skills in High-, Average-, and Low-Achieving Primary School Children. *Journal of Reading Behavior*, 26(4), 361-389.
- Vogelsang, C. & Reinhold, P. (2013). Gemessene Kompetenz und Unterrichtsqualität. Überprüfung der Validität eines Kompetenztests mit Hilfe der Unterrichtsvideographie. In U. Riegel & K. Macha (Hrsg.), *Videobasierte Kompetenzforschung in den Fachdidaktiken. Fachdidaktische Forschungen (Band 4)*. Münster: Waxmann.
- von Aufschnaiter, C., Cappell, J., Dübbelde, G., Ennemoser, M., Mayer, J., Stiensmeier-Pelster, J. et al. (2015). Diagnostische Kompetenz - Theoretische Überlegungen zu einem zentralen Konstrukt der Lehrerbildung. *Zeitschrift für Pädagogik*, 61(5), 738-758.
- Walter, J. (2010). *LDL. Lernfortschrittsdiagnostik Lesen. Ein curriculumbasiertes Verfahren / Manual*. Göttingen: Hogrefe.
- Walter, J. (2013). *VSL. Verlaufsdiagnostik sinnerfassenden Lesens / Manual*. Göttingen: Hogrefe.
- Warnke, A., Hemminger, U. & Plume, E. (2004). *Lese-Rechtschreibstörungen*. Göttingen: Hogrefe.
- Warnke, A., Hemminger, U., Roth, E. & Schneek, S. (2002). *Legasthenie. Leitfaden für die Praxis. Begriff – Erklärung – Diagnose – Behandlung – Begutachtung*. Göttingen: Hogrefe.
- Wehr, S. & Hartmann, E. (2005). Anforderungen und Hürden im frühen Schriftspracherwerb. Was Lehrpersonen über einen entwicklungsgerechten Anfangsunterricht wissen sollten. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete (VHN)*, 74(3), 196-206.
- Weinert, F. E. (2000). Lehren und Lernen für die Zukunft – Ansprüche an das Lernen in der Schule. *Pädagogische Nachrichten Rheinland-Pfalz*(2), 1-16.
- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17-31). Weinheim: Beltz.
- Wember, F. B. (1999). *Besser lesen mit System*. Neuwied: Luchterhand.
- Wember, F. B. (2015). Weiterführendes Lesen. In U. Heimlich & F. B. Wember (Hrsg.), *Didaktik des Unterrichts im Förderschwerpunkt Lernen. Ein Handbuch für Studium und Praxis* (3. Aufl., S. 191-205). Stuttgart: Kohlhammer.
- Wendt, H. & Bos, W. (2011). Fachdidaktik und Bildungsforschung – von der Notwendigkeit zur Kooperation im Zeitalter globalisierter Kompetenzen. In K.-O. Bauer & N. Logemann (Hrsg.), *Unterrichtsqualität und fachdidaktische Forschung. Modelle und Instrumente zur Messung fachspezifischer Lernbedingungen und Kompetenzen* (S. 11-34). Münster: Waxmann.

- Wieser, D. (2008). *Literaturunterricht aus Sicht der Lehrenden. Eine qualitative Interviewstudie*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Wiesner, E. & Schneider, H. (2011). Wenn Lesen trotzdem gelingt. In H. Schneider (Hrsg.), *Wenn Schriftaneignung (trotzdem) gelingt. Literale Sozialisation und Sinneserfahrung* (S. 157-244). Weinheim: Beltz, Juventa.
- Wilbert, J. & Linnemann, M. (2011). Kriterien zur Analyse eines Tests zur Lernverlaufsdiagnostik. *Empirische Sonderpädagogik*, 3(3), 225-242.
- Wild, J. & Pissarek, M. (2018). *RATTE: Regensburger Analysetool für Texte Dokumentation*. Verfügbar unter: <http://www.uni-regensburg.de/sprache-literatur-kultur/germanistik-did/ratte/index.html> [Zugriff: Oktober 2019].
- Wimmer, H. (1993). Characteristics of Developmental Dyslexia in a Regular Writing System. *Applied Psycholinguistics*, 14(1), 1-33.
- Wimmer, H. & Hummer, P. (1990). How German-Speaking First Graders Read and Spell: Doubts on the Importance of the Logographic Stage. *Applied Psycholinguistics*, 11(4), 349-368.
- Wimmer, H., Landerl, K., Linortner, R. & Hummer, P. (1991). The relationship of phonemic awareness to reading acquisition: More consequence than precondition but still important. *Cognition* 30(3), 219-249.
- Winkes, J. (2014). *Isolierte Rechtschreibstörung. Eigenständiges Störungsbild oder leichte Form der Lese- Rechtschreibstörung? Eine Untersuchung der kognitiv-linguistischen Informationsverarbeitungskompetenzen von Kindern mit Schriftspracherwerbsstörungen*. [Dissertation der philosophischen Fakultät der Universität Freiburg/CH]. Verfügbar unter: <https://edudoc.ch/record/116080?ln=de> [Zugriff: Juni 2020].
- Winther, E. (2010). *Kompetenzmessung in der beruflichen Bildung*. Bielefeld: Bertelsmann.
- Wiprächtiger-Geppert, M., Riegler, S. & Freivogel, J. (2015). Erfassung des professionellen Wissens von Deutschlehrkräften zu Orthographie und Orthographieerwerb – Forschungsstand und Perspektiven. In C. Bräuer & D. Wieser (Hrsg.), *Lehrende im Blick: Empirische Lehrerforschung in der Deutschdidaktik* (S. 281-300). Wiesbaden: Springer VS.
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.
- Wittmann, G. (2017). Wissen und Überzeugungen – zentrale Aspekte der mathematikbezogenen professionellen Kompetenz pädagogischer Fachkräfte. In S. Schuler, C. Streit & G. Wittmann (Hrsg.), *Perspektiven mathematischer Bildung im Übergang vom Kindergarten zur Grundschule* (S. 189-206). Wiesbaden: Springer.
- Wolf, M. & Bowers, P. (1999). The double-deficit hypothesis for the developmental dyslexia Journal of Educational Psychology. *Journal of Educational Psychology*, 91(3), 415-438.
- Wu, M. (2005). The Role of Plausible Values in Large-Scale Surveys. *Studies in Educational Evaluation*, 31(2), 114-128.
- Wu, M. & Adams, R. J. (2013). Properties of Rasch residual Fit Statistics. *Journal of applied measurement*, 14(4), 339-255.
- Wu, M., Adams, R. J. & Wilson, M. R. (2007, Version 2.0). *ACER ConQuest / Manual*. Australia: ACER Press.
- Zimmermann, H. & Peyer, A. (2016). Wissen und Normen – Facetten professioneller Kompetenz von Deutschlehrkräften. Eine Einleitung zum Tagungsband. In H. Zimmermann & A. Peyer (Hrsg.), *Wissen und Normen – Facetten professioneller Kompetenz von Deutschlehrkräften* (S. 7-21). Frankfurt am Main: Peter Lang.
- Zlatkin-Troitschanskaia, O. & Kuhn, C. (2010). Lehrprofessionalität. Ein Überblick zum theoretischen und methodischen Stand der nationalen und internationalen Forschung. In J. König & B. Hofmann (Hrsg.), *Professionalität von Lehrkräften – Was sollen Lehrkräfte im Lese- und Schreibunterricht wissen und können?* (S. 24-40). Berlin: Deutsche Gesellschaft für Lesen und Schreiben.
- Zumbo, B. D. (2007). Three Generations of DIF Analyses: Considering Where It Has Been, Where It Is Now, and Where It Is Going. *Language Assessment Quarterly*, 4(2), 223-233.

10 Abbildungsverzeichnis

Abbildung 1. Mehrebenenmodell des Lesens (Rosebrock & Nix, 2012, S. 11).	23
Abbildung 2. Zwei-Wege-Modell des Wortlesens (Scheerer-Neumann, 1998, S. 15)	25
Abbildung 3. Das Zweidimensionale Modell der phonologischen Bewusstheit von Schnitzler (2008) mit Aufgabenbeispielen von Mayer (2013, S. 49).	46
Abbildung 4. Kompetenzentwicklungsmodell für das Worterkennen und laute Lesen (Klicpera et al., 2017, S. 29).....	51
Abbildung 5. Formel für die Berechnung des Lix (Wild & Pissarek, 2018, S. 9)	58
Abbildung 6. Entwicklung der Lesegeschwindigkeit (Wörter pro Minute) von der zweiten bis zur achten Klasse von Kindern mit anfänglich unterschiedlichen Lesefähigkeiten (entnommen aus Schneider, 2017, S. 83)	65
Abbildung 7. Kompetenz als Kontinuum (Blömeke et al., 2015a, S. 7).....	86
Abbildung 8. Modell professioneller Kompetenz bei COACTIV mit Spezifikationen für das Professionswissen	95
Abbildung 9. Wirkkette der schulischen Bildung aus Vogelsang und Reinhold (2013, S. 319) mit gendergerechter Anpassung durch Jandl (2016, S. 39).....	118
Abbildung 10. Beispiel-Item 8 Fachwissen (Kennen/Erinnern/Abrufen).....	132
Abbildung 11. Beispiel-Item 26 fachdidaktisches Wissen/Können (Verstehen/Analysieren/Anwenden).....	134
Abbildung 12. Untersuchungsplan und Arbeitsschritte.....	142
Abbildung 13. Beispiel-Item 28 mit offenem Kurzantwort-Format zum Lesbarkeitsindex.....	148
Abbildung 14. Beispiel-Item 5, Single-Choice mit einer korrekten Antwortmöglichkeit (max. 5 Punkte).....	155
Abbildung 15. Beispiel-Item 3, Audio-Item mit mehreren Antwortmöglichkeiten (max. 6 Punkte)	156
Abbildung 16. Beispiel-Item 28 mit offenem Antwortformat.....	157
Abbildung 17. Exemplarische Itemcharakteristikfunktion (ICC) im dichotomen Rasch-Modell, ergänzte Graphik.....	167
Abbildung 18. Beispiel einer Verteilung der Itemschwierigkeit in Relation zur Personenfähigkeit der latenten Dimension. ConQuest Manual V2 (Wu, Adams & Wilson, 2007, Version 2.0)	171
Abbildung 19. Theoretische Modelle, gegliedert nach inhaltlichen Dimensionen und kognitiven Anforderungen	178
Abbildung 20. Map of latent Distribution: Verteilung der Itemschwierigkeit (rechts) in Relation zur Verteilung der Personenfähigkeit (links) in der Voruntersuchung.....	185
Abbildung 21. Beispiel Audio-Item 1 vor und nach der Überarbeitung in der Voruntersuchung und Hauptstudie	186
Abbildung 22. Map of latent Distribution: Gegenüberstellung der Verteilung der Itemschwierigkeit (rechts) in Relation zur Personenfähigkeit (links) im Rasch-Modell in der Skala ₂₈ ..	195
Abbildung 23. Map of latent Distribution: Verteilung der Itemschwierigkeiten (rechts) in Relation zur Verteilung der Personenfähigkeit (links) in der Hauptuntersuchung Skala ₁₈ und Berücksichtigung der Itemformate	196
Abbildung 24. Itemschwierigkeit der Substichprobe I, II und III im Vergleich bei gemeinsamer fixierter Skala (Skala ₁₈).....	197
Abbildung 25: Beispiel-Item 18 mit hoher Wissensanforderungen (offenes Antwortformat)	198
Abbildung 26. Beispiel-Item 19 mit hoher Wissensanforderungen (gebundenes Antwortformat) ..	198
Abbildung 27. Beispiel-Item 5 mit niedriger Wissensanforderungen (gebundenes Antwortformat)	199
Abbildung 28. Beispiel-Item 2 (Audio-Item) mit unplausibler Kategorisierung bei Distraktor f....	202
Abbildung 29. Beispiel-Item 5 mit unplausiblen Distraktoren a, b, d	203
Abbildung 30. Vergleich der drei Substichproben hinsichtlich ihrer relativen Abweichung der Itemschwierigkeit vom Mittel wert Null in Skala ₁₈	204

Abbildung 31. Streudiagramm zum Vergleich der relativen Itemschwierigkeit von Substichprobe Primar _{ohneBE} und Substichprobe SHP	205
Abbildung 32. Beispiel-Item 22 zur Analyse von Anlauttabellen.....	207
Abbildung 33. Beispiel-Item 6 mit inhaltlicher Nähe zu Berufserfahrungswissen.....	208
Abbildung 34. Beispiel-Item 18 mit inhaltlicher Nähe zu Berufserfahrungswissen.....	208
Abbildung 35. Beispiel-Item 17 zum Leseverständnis.....	209
Abbildung 36. Beispiel-Item 19 mit DIF in allen drei Substichproben.....	210
Abbildung 37. Scree-Test zur Extraktion von Faktoren auf Grundlage des Eigenwertkriteriums mit Skala ₁₈	215
Abbildung 38. Beispiel-Item 14 für ausbildungsspezifische Wissensanforderungen	220
Abbildung 39. Beispiel-Item 28 für ausbildungsspezifische Wissensanforderungen	220
Abbildung 40. Vergleich des querschnittlichen Wissensanstiegs auf Grundlage der Personenfähigkeits-Schätzer der Substichproben und Kontraststichprobe (N = 507).	223
Abbildung 41. Scree-Test zur Extraktion von Faktoren auf Grundlage des Eigenwertkriteriums aus der Voruntersuchung.....	286
Abbildung 42. Verwendete Formel zur statischen Absicherung der DIF von Draba (1977)	286

11 Tabellenverzeichnis

Tabelle 1: Adaptierte Darstellung der Entwicklungsstadien Lehrerexpertise nach Berliner (2001), abgedruckt in König (2010, S. 53)	88
Tabelle 2: Ausformulierte Items und Lösungswahrscheinlichkeiten für die Skala Grundwissen zum Lesen/LRS aus (Schmidt & Schabmann, 2016).....	103
Tabelle 3: Modell des lesebezogenen Wissens/Könnens	129
Tabelle 4: Verortung der Items im Modell des lesebezogenen Wissens/Könnens	145
Tabelle 5: Theoriebezug der Items zu Facetten des Grundlagenwissens und diagnostischen Wissens	146
Tabelle 6: Verteilung der Itemformate im Modell in der Hauptstudie	151
Tabelle 7: Übereinstimmungsreliabilität bei Items mit offenem Format in der Hauptuntersuchung	158
Tabelle 8: Inhaltliche Beschreibung der Substichproben nach Kriterien der Berufsbiografie	161
Tabelle 9: Quantitative Beschreibung der Substichproben in der Hauptstudie	162
Tabelle 10: Substichproben und Übersicht der durchgeführten Analysen der Voruntersuchung.....	183
Tabelle 11: Statistische Kennwerte als Grundlage für die Itemanalyse in der Voruntersuchung mit 15 Items	184
Tabelle 12: Dimensionalität des Befragungsinstruments in der Voruntersuchung.....	187
Tabelle 13: Gruppenunterschiede zwischen Substichproben und Laien-Kontraststichprobe, Bonferroni Test.....	188
Tabelle 14: Substichproben und Übersicht der durchgeführten Analysen der Hauptstudie.....	189
Tabelle 15: Statistische Kennwerte als Grundlage für die Itemanalyse in der Hauptuntersuchung.	190
Tabelle 16: Mittelwert, Standardabweichung, Prozentualer Anteil erreichter Punkte, Trennschärfe und Reliabilität pro Substichproben (Skala ₁₈)	191
Tabelle 17: Weighted In-Fit (MNSQ) für Skala ₂₈ und Skala ₁₈	192
Tabelle 18: Weighted In-Fit (MNSQ) getrennt nach Substichproben (Skala ₁₈)	193
Tabelle 19: Gütekriterien auf Grundlage des Rasch-Modells für Skala ₂₈ und Skala ₁₈	193
Tabelle 20: Antwortverhalten Beispiel-Item 19 getrennt nach Substichproben	199
Tabelle 21: Antwortverhalten Beispiel-Item 5 getrennt nach Substichproben	199
Tabelle 22: Analyse zur Beurteilung der Plausibilität von Distraktoren	201
Tabelle 23: Differenzielle Itemfunktion in den drei Substichproben im paarweisen Vergleich	206
Tabelle 24: Skalenvergleich unter Berücksichtigung von DIF auf Grundlage der WLE-Mittelwerte (Skala ₁₈).....	211
Tabelle 25: Modellvergleich Grundlagenwissen und diagnostisches Wissen vs. eindimensionales Modell für Skala ₂₈ und	213
Tabelle 26: Modellvergleich Fachwissen und fachdidaktisches Können vs. eindimensionales Modell für Skala ₁₈ und Skala ₂₈	214
Tabelle 27: Eigenwerte, Prozent der erklären Varianz und kumulierte Erklärte Varianz der fünf extrahierten Faktoren, Skala ₁₈	216
Tabelle 28: Vergleich von Faktorenlösungen im Rahmen der exploratorischen Faktorenanalyse...	216
Tabelle 29: Verortung der Items im zweidimensionalen Modell	218
Tabelle 30: Gruppenunterschiede zwischen Substichproben und Laien-Kontraststichprobe, Games- Howell-Test	224
Tabelle 31: Übersicht der Werte auf Itemebene der Skala ₁₈ nach KTT und IRT (Rasch-Modell) ...	226
Tabelle 32: Übersicht der Werte auf Skalenebene der Skala ₁₈ nach KTT und IRT	227
Tabelle 33: Übersicht der exploratorischen Faktorenanalyse, eindimensionales und zweidimensionales Modell.....	228
Tabelle 34: Mittelwerte der fünf Substichproben	287
Tabelle 35: Differentielle Itemfunktionen: Vergleich von I Primar _{ohneBE} und II Primar _{mitBE}	287
Tabelle 36: Differentielle Itemfunktionen: Vergleich von II Primar _{mitBE} III SHP	287
Tabelle 37: Differentielle Itemfunktionen: Vergleich von I Primar _{ohneBE} und III SHP	288
Tabelle 38: Faktorladungen und Kommunalitäten des eindimensionalen Modells in der Hauptuntersuchung	288
Tabelle 39: Reduziertes Modell des lesebezogenen Wissens/Könnens und Verortung der Items ...	289

12 Anhang

In der Arbeit erwähnte Abbildungen

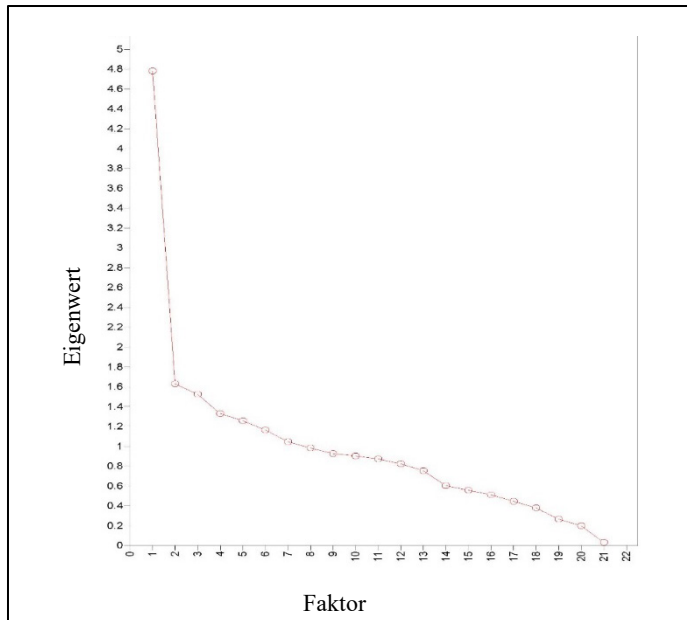


Abbildung 41. Scree-Test zur Extraktion von Faktoren auf Grundlage des Eigenwertkriteriums aus der Voruntersuchung

$$t_{(1,2)} = \frac{|d_{(1)} - d_{(2)}|}{\sqrt{Se_{(1)}^2 + Se_{(2)}^2}} > 1.96$$

$d_{(1)}$ = Logit Gruppe 1

$d_{(2)}$ = Logit Gruppe 2

$Se_{(1)}$ = Standardfehler Gruppe 1

$Se_{(2)}$ = Standardfehler Gruppe 2

Abbildung 42. Verwendete Formel zur statischen Absicherung der DIF von Draba (1977)

In der Arbeit erwähnte Tabellen

Tabelle 34: Mittelwerte der fünf Substichproben

K	Substichprobe	n	M	SD	Punkte Min.	Punkte Max.
I	Primarlehrpersonen ohne Berufserfahrung am Ende der Ausbildung	190	19.43	4.24	10	29
II	Primarlehrpersonen mit Berufserfahrung	127	20.71	3.56	11	29
III	SHP mit Berufserfahrung	105	26.70	5.11	8	36
IV	SHP ohne Berufserfahrung	9	24.11	5.51	13	33
V	Zwischenstufen / fachfremde Grundausbildung	12	23.50	5.70	15	31
Σ		443				

Anmerkung. Die Mittelwerte wurden aufgrund der erreichten Punkte unter Einbezug von 28 Items berechnet. Min. und Max. bezieht sich auf den Range bei Total 42 möglichen Punkten

Tabelle 35: Differentielle Itemfunktionen: Vergleich von I Primar_{ohneBE} und II Primar_{mitBE}

Item Nr.	I Primar _{ohneBE}		II Primar _{mitBE}		Betrag Differenz der Logit-Werte	t	p	Bewertung DIF
	Logit	SE	Logit	SE				
2	-1.242	0.130	-0.986	0.160	0.26	1.37	0.210	vernachlässigbar
5	-1.473	0.164	-2.078	0.247	0.60	2.04	0.042	moderat
6	-0.962	0.152	-1.444	0.205	0.48	1.89	0.060	moderat
7	0.546	0.127	0.601	0.133	0.06	0.31	0.755	vernachlässigbar
8	0.701	0.140	0.268	0.128	0.43	2.28	0.023	moderat
9	-0.545	0.148	-0.019	0.172	0.53	2.32	0.020	moderat
11	-0.017	0.151	0.072	0.172	0.09	0.39	0.689	vernachlässigbar
12	-1.530	0.165	-1.490	0.207	0.04	0.16	0.873	vernachlässigbar
13	-0.213	0.096	0.037	0.108	0.25	1.72	0.081	vernachlässigbar
14	1.211	0.190	1.438	0.215	0.23	0.81	0.424	vernachlässigbar
16	-1.655	0.135	-1.809	0.228	0.15	0.59	0.567	vernachlässigbar
17	-1.588	0.167	-0.938	0.184	0.65	2.62	0.009	hoch
18	2.229	0.272	1.439	0.215	0.79	2.26	0.023	hoch
19	2.314	0.281	3.530	0.484	1.22	2.19	0.030	hoch
22	0.079	0.114	-0.045	0.131	0.12	0.71	0.484	vernachlässigbar
23	-0.040	0.151	-0.799	0.181	0.76	3.22	0.001	hoch
27	0.604	0.124	0.806	0.141	0.20	1.09	0.276	vernachlässigbar
28	1.816	0.272	1.608	0.211	0.21	0.57	0.550	vernachlässigbar

Tabelle 36: Differentielle Itemfunktionen: Vergleich von II Primar_{mitBE} III SHP

Item Nr.	II Primar _{mitBE}		III SHP		Betrag Differenz der Logit-Werte	t	p	Bewertung DIF
	Logit	SE	Logit	SE				
2	-0.986	0.160	-0.824	0.241	0.17	0.60	0.551	vernachlässigbar
5	-2.078	0.247	-1.684	0.328	0.41	0.99	0.324	vernachlässigbar
6	-1.444	0.205	-0.692	0.243	0.76	2.40	0.016	hoch
7	0.601	0.133	0.308	0.148	0.28	1.42	0.156	vernachlässigbar
8	0.268	0.128	0.874	0.153	0.62	3.09**	0.002	moderat
9	-0.019	0.172	0.275	0.206	0.31	1.14	0.256	vernachlässigbar
11	0.072	0.172	-0.292	0.223	0.35	1.26	0.210	vernachlässigbar
12	-1.490	0.207	-1.679	0.328	0.18	0.46	0.646	vernachlässigbar
13	0.037	0.108	-0.366	0.170	0.39	1.95	0.051	vernachlässigbar
14	1.438	0.215	0.448	0.203	0.98	3.31**	0.001	hoch
16	-1.809	0.228	-2.006	0.495	0.19	0.34	0.732	vernachlässigbar
17	-0.938	0.184	-0.631	0.240	0.32	1.05	0.294	vernachlässigbar
18	1.439	0.215	2.581	0.249	1.15	3.50**	0.000	hoch
19	3.530	0.484	0.980	0.200	2.55	4.87**	0.000	hoch
22	-0.045	0.131	0.583	0.160	0.64	3.09**	0.002	moderat
23	-0.799	0.181	0.140	0.209	0.95	3.44**	0.001	hoch
27	0.806	0.141	1.170	0.152	0.38	1.80	0.071	vernachlässigbar
28	1.608	0.211	0.815	0.149	0.78	3.03**	0.002	hoch

Tabelle 37: Differentielle Itemfunktionen: Vergleich von I Primar_{ohneBE} und III SHP

I Primar _{ohneBE}			III SHP					
Item					Betrag	Differenz		
Nr.	Logit	SE	Logit	SE	der Logit-Werte	t	p	Bewertung DIF
2	-1.242	0.130	-0.824	0.241	0.43	1.57	0.115	vernachlässigbar
5	-1.473	0.164	-1.684	0.328	0.20	0.54	0.589	vernachlässigbar
6	-0.962	0.152	-0.692	0.243	0.28	0.99	0.323	vernachlässigbar
7	0.546	0.127	0.308	0.148	0.26	1.15	0.249	vernachlässigbar
8	0.701	0.140	0.874	0.153	0.19	0.90	0.370	vernachlässigbar
9	-0.545	0.148	0.275	0.206	0.83	3.28	0.001	hoch
11	-0.017	0.151	-0.292	0.223	0.26	0.97	0.331	vernachlässigbar
12	-1.530	0.165	-1.679	0.328	0.14	0.37	0.711	vernachlässigbar
13	-0.213	0.096	-0.366	0.170	0.14	0.72	0.473	vernachlässigbar
14	1.211	0.190	0.448	0.203	0.75	2.69	0.007	hoch
16	-1.655	0.135	-2.006	0.495	0.34	0.66	0.510	vernachlässigbar
17	-1.588	0.167	-0.631	0.240	0.97	3.32	0.001	hoch
18	2.229	0.272	2.581	0.249	0.37	1.99	0.322	vernachlässigbar
19	2.314	0.281	0.980	0.200	1.33	3.87	0.000	hoch
22	0.079	0.114	0.583	0.160	0.52	2.63	0.008	moderat
23	-0.040	0.151	0.140	0.209	0.19	0.75	0.454	vernachlässigbar
27	0.604	0.124	1.170	0.152	0.58	2.95	0.003	moderat
28	1.816	0.272	0.815	0.149	0.99	3.19	0.001	hoch

Tabelle 38: Faktorladungen und Kommunalitäten des eindimensionalen Modells in der Hauptuntersuchung

Item Nr.	Faktorladungen λ	Kommunalität h^2
2	.30*	.09
5	.51*	.27
6	.40*	.16
7	.61*	.38
8	.50*	.25
9	.32*	.10
11	.41*	.17
12	.57*	.33
13	.59*	.35
14	.65*	.42
16	.56*	.32
17	.35*	.13
18	.39*	.15
19	.76*	.58
22	.29*	.08
23	.32*	.10
27	.54*	.29
28	.89*	.80
Eigenwert	5.43	
% der erklärten Varianz	30.17	

* < .05

Tabelle 39: Reduziertes Modell des lesebezogenen Wissens/Könnens und Verortung der Items

Inhaltsbereich	Kognitive Anforderung			
	Item Nr.	Fachwissen Kennen/Erinnern/Abrufen	Item Nr.	Fachdidaktisches Wissen/Können Verstehen/Analysieren/Anwenden
<i>Facette 1</i>				
Wissen über linguistische Merkmale der deutschen Schriftsprache und Aspekte der Textschwierigkeit	28	Was ist der Lix und über welche Textmerkmale wird er berechnet	26 27	Ordnen der Satzschwierigkeit Phonemanalyse von 6 Begriffen
<i>Facette 2</i>				
Wissen über Prozessmodelle des Lesens und Determinanten der Lesekompetenz	9 10 19	Zusammenhang Leseflüssigkeit und Leseverständnis Zusammenhang Hörverständnis Leseverständnis Mindestanteil an korrekt gelesenen Wörtern für unabhängiges Lesen		
<i>Facette 3</i>				
Wissen über Stufenmodelle von Leseerwerb / Leseentwicklung inkl. Vorläuferfertigkeiten	8	Identifizieren von Elementen der phonologischen Bewusstheit	6 7	Identifizieren der einfachsten Leseaufgabe auf Wortebene Ordnen von Leseaufgaben nach Lesestrategie
<i>Facette 4</i>				
Wissen über didaktische Konzepte und Hilfestellungen für den Leseunterricht	17 22	Vermittlung von Lesestrategien für das Leseverständnis Analyse von Beispielen bei Anlauttabelle	21 23 24 25	Wirksamkeit von Stilleseverfahren Hilfestellung zur Leseflüssigkeit mit Silbenbogen Maßnahmen zur Förderung Lesestrategie Wissen über effektive Förderung der Leseflüssigkeit
Σ 7 Items			Σ 8 Items	
<i>Facette 1</i>				
Wissen über Symptome von Leseschwierigkeiten	11 15	Typische Leseschwierigkeiten im Anfangsunterricht Typische Leseschwierigkeiten 3. Klasse	1 2	Analyse von Lesekompetenzen 2. Klasse Analyse von Lesekompetenzen 4. Klasse
<i>Facette 2</i>				
Wissen über Leseverläufe bei unauffällig und schwach lesenden Kindern / Wissen über typische Schwierigkeiten bei Stufenübergängen	12	Typische Symptome von Leseschwierigkeiten im Vergleich zu gut lesenden Kindern	3 4 5	Merkmale Stufe der Leseentwicklung Diagnose und Förderung LRS 2. Klasse Identifizieren einer Leseschwierigkeit auf Wortebene
<i>Facette 3</i>				
Wissen über diagnostische Verfahren und Aufgaben zur Erfassung von Kompetenzen zu Teilbereichen des Lesens	20	Diagnose von ersten lesebezogenen Auffälligkeiten	13 14 16 18	Diagnose mittels 1-Minuten-Leseprobe Analyse einer Aufgabe zur Benennungsgeschwindigkeit Spezifisches, diagnostisches Potenzial von Leseaufgaben Entwickeln einer Leseaufgabe für Kinder, die ratend lesen
Σ 4 Items			Σ 9 Items	